Profiling Complex Online Interactions: What Behaviour Can You Infer From a Digital Footprint?

Mohamed Mostafa and **Tom Crick** and **Giles Oatley**
Department of Computing & Information Systems
Cardiff Metropolitan University
Cardiff, UK

Abstract

This is an initial exploration of how people interact with complex online information systems, using a wide range of digital data and employing a number of personality and behavioural representation systems. The information system is an online application portal for an international academic mobility grant scheme, and we analyse document uploading behaviour, motivation letters and social media interactions.

We apply psycholinguistic techniques to the motivation letters to determine the Five Factors personality scorings, and apply the same techniques to interactions with a dedicated Facebook page. The relative rankings are compared using the Kendall rank correlation statistic.

Finally, we examine the upload footprint for the users and determine several classes of behaviour. These again are compared against the Five Factors, and in turn against the eventual grant status of the applicant.

Introduction

Understanding how to develop software to facilitate its effective and efficient use is a core part of the broad field of human-computer interaction. Different users form different conceptual models about their interactions and have different ways of obtaining and developing knowledge and skills; cultural and national differences may also play a significant role. Another consideration in human-computer interaction is that technology – and in particular, user interface technology – changes rapidly, offering new interaction possibilities to which previous research findings may not necessarily apply. Alongside this, user preferences (and the way in which they interact with the software) change as they gradually master new interfaces and environments.

Personality and behaviour can be determined from digital data (Pennebaker, Francis, and Booth 2001; Vazire and Gosling 2004; Iacobelli et al. 2011; Oatley and Crick 2014); while in the past this has mainly been the textual information contained in blogs, status posts and photo comments (Blamey, Crick, and Oatley 2012; 2013), there is also a wealth of information in the other ways of interacting with digital artefacts. For instance, it is possible to observe the ordering (and frequency) of button clicks for a user.

Mairesse et al. (2007) demonstrate the use of features from the psycholinguistic databases LIWC (Pennebaker, Francis, and Booth 2001) and MRC (Wilson 1988) to create a range of statistical models for each of the Five Factors personality traits (Norman 1963; Peabody and Goldberg 1989; Goldberg 1990). These five traits are: *Extraversion*, *Emotional Stability*, *Agreeableness*, *Conscientiousness* and *Openness to Experience*. Equation 1 describes *Extraversion*, where each feature is prefixed by the containing database.

```
Extraversion =
            -0.0379 * MRC.K_F_NSAMP +
            -0.0803 * LIWC.UNIQUE +
            -0.6074 * LIWC.ABBREVIATIONS +
             0.1445 * LIWC.PRONOUN +
            -0.3941 * LIWC.HEARING +
            17.1407;
```
(1)

The test system analysed in this paper is the user interaction for a web-based educational application portal: "*Online Portal for Scholarship Mobility*". We make use of textual data, analysing with these same psycholinguistic techniques, and employ standard statistical methods on non-textual data. The textual data also includes interaction with a dedicated Facebook page dedicated to resolving problems with the applications (of which there were many); and, the actual documents submitted, including a free-text application motivation letter. The non-textual data includes the final scoring of the individual for the grant they applied for (e.g. success, reserved, failure) and the individual's behaviour on the site (when they uploaded their documents, how close to the deadline and so on).

Domain Description and Available Data

Our data comes from an online portal for a European Union (EU) international scholarship mobility hosted at a UK university. The aim of the mobility programme is to enhance quality in higher education through scholarships and academic cooperation between Europe and the rest of the world. It provides mobility grants for students at different academic levels (Undergraduate, Masters, PhD, Post-Doctoral, Faculty) and has numerous courses available from a wide range of institutions across the EU.

The details of the call were as follows: there were 2,706 applications submitted by 1,170 candidates, applying to 10 EU universities and 10 non-EU universities. The system allows an applicant to apply to up to three courses from all courses offered by the 10 universities, and the applicant is required to assign a priority for each course. This priority field is used in the selection stage, for instance if the applicant assigned Course A as priority (1) and Course B as priority (2) and the applicant is accepted in both courses, then the first priority will be offered.

Each mobility call has an opening date/time and closing date/time, with occasional extensions given for specific reasons (for instance due to administrative reasons or technical issues with the portal). Applicants are required to submit for their application certain mandatory files, such as motivation letter, passport/identification, curriculum vitae), as well as optional files (supporting documents). The primary modes of communication between candidates and the project team is via emails, telephone and the dedicated Facebook page. The selection process divided into three stages: Eligibility, Evaluation and Final Selection.

Eligibility

The project team scans each document to make sure all documents are uploaded correctly, for example users might upload an empty image in place of motivation letters and these need to be manually verified. There are mandatory documents to be uploaded to meet the minimum requirements of the call.

Evaluation

Eligible users are evaluated by a technical expert from the host university, based on the following: motivation letter; academic merit/performance (qualifications/grades); English language competencies e.g. International English Language Testing System (IELTS)[1] score; and academic/research proposal.

Final Selection

Based on which rank the applicant assigned to the host university, the final selection is the top *n* of applicants. *n* is calculated based in the host capacity and budget of the project. This process results in the final classification of the applicant as either:

- *Accepted* (ranked highest);
- *Reserved* (passed but not selected);
- *Rejected* (below passing grade);
- *Ineligible* (missing documents or out-dated documents).

Experimental Design

Initially, we carried out three different types of experiments: Experiment 1 compared the motivation letters against the Facebook interactions; Experiment 2 compared the interaction footprint against the motivation letters; Experiment 3 checks the raw data using multiple regression.

[1] `http://www.ielts.org/`

Experiment 1

Applicants are required to upload a description of why they are applying for this particular mobility grant, the motivation letter. Applicants also communicated with the project team through the project Facebook page.

We are interested in the similarity between language characteristics used in the much more formal documents (for instance the motivation letter) and that used in online social media (the Facebook page). The Facebook interaction consists of posts, comments and "likes". We extracted the text from all motivation letters and Facebook interactions and analysed both blocks of text according to the Five Factor personality traits as discussed previously.

After extracting the Five Factors for each applicant we compared the relative positions of each of the Five Factors in both lists using the Kendall rank correlation statistic (Kendall 1938). We did this for: all applicants (*All*); only the accepted applicants (*Accepted*); only the reserved applicants (*Reserved*); and only the rejected applicants (*Rejected*).

The average Kendall's tau coefficient value is reported for these groups, for each of the Five Factor features. By considering rank position and not absolute value, we mitigate against explaining values without baseline experimentation.

Experiment 2

We simplified an applicant's interaction, or timeline, with the portal to include the following milestones: *T0* Registration Time; *T1* First Action; *T2* Last Action; and, *T3* Submission. Additionally we represented an extension to the submission deadline as *T4* Extension. In this way we can represent an applicant's interaction as shown in Figure 1, which shows seven example timelines.

Using these milestones we are able to identify interesting behaviours that compare and contract with personality traits and other sources of information. Behaviours such as: how long it was before an applicant became aware of the call, and when they registered; how long after registration did the applicant carry out their first action with the system; how long did they take to complete their application; and, how close to the deadline did they submit their application.

We divided the timeline of the call into five segments as presented in the following Table 1. The complete timeline from opening to final close was 125 days. There was an extension from day 112 until day 125. The segments or timeline periods are determined as percentage chunks of the total timeline, for instance segment *S0* is the first 20% of the timeline, and so ranges from day 1 until day 25, segment *S1* ranges from day 26 until day 50, and so on.

Segment	Start	Finish
S0	0	20
S1	20	40
S2	40	60
S3	60	90
S4	90	100

Table 1: Timeline periods as percentages of total timeline

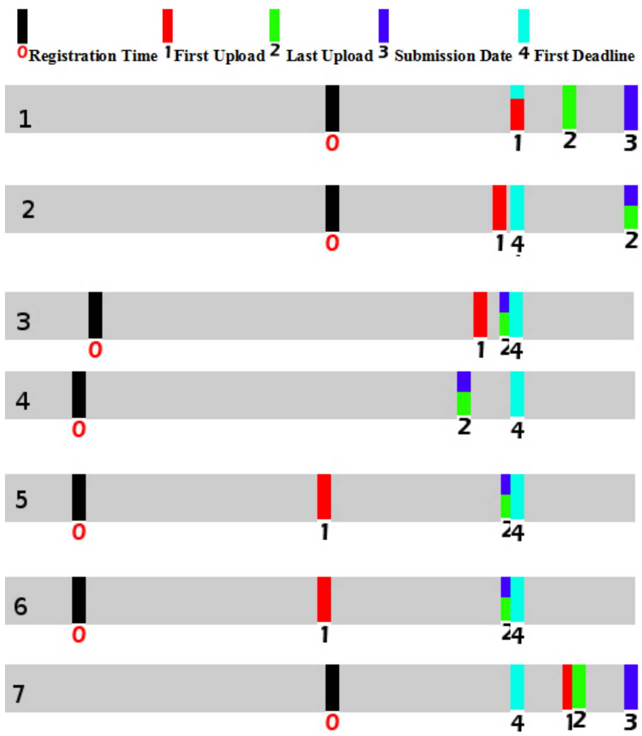

Figure 1: Seven user timelines. *T0* (black bar) is when the applicant first registered with the call. *T1* (red bar) represents when the applicant uploaded their first document, or First Action. *T2* (green bar) represents an applicants' Last Action. *T3* (blue bar) represents the applicants' Submission. *T4* (aquamarine bar) represents the first deadline (certain calls had initial deadlines extended)

Using these segments we were able to assign the various applicant actions (*T0* Registration, *T1* First Upload, *T2* Last Upload, *T3* Submission) to various time periods. This allowed us to assign applicants to statistically significant categories, and also to add in a few categories from observations. These are shown in Table 2; as you can see, a small number of applicants (*n*=4) registered within the segment S1 (20-40% of timeline), and then uploaded all of their documents and submitted within the segment S3 (60-90% of timeline). This is represented by Class A, the first row; successive rows can be interpreted in the same manner.

Class	*n*	*T0*	*T1*	*T2*	*T3*
A	4	S1	S3	S3	S3
B	14	S2	S2	S2	S2
C	128	S2	S3	S3	S3
D	29	S2	S3	S4	S4
E	678	S3	S3	S3	S3
F	202	S3	S3	S4	S4
G	9	S3	S4	S4	S4
H	54	S4	S4	S4	S4

Table 2: Applicants' timeline actions assigned to segments

We do not want to be too quick to ascribe an alias to the behaviours, as we recognise that there are several possible interpretations; nevertheless, we have used the 'Potential Alias' column in Table 3 to indicate some initial thoughts.

Experiment 3

While the compound features of the Five Factors are an interesting perspective, we also needed to check the raw data underneath this, in the form of the psycholinguistic features LIWC and MRC. For this investigation we chose multiple regression.

We extracted the LIWC row data features (87 features) from the motivation letters and analysed the input data set against the 'status' of the application. The method used was multiple regression, the dependent variable being 'status' and the independent variables are the LIWC features. We proceeded as follows:

- Our first assumption is multicollinearity, which refers to the relationship when two independents variables are highly correlated;
- Removing the above features, carry out regression;
- Detecting outliners using Mahalanobis distance, and since we have 60 features remaining after the multicollinearity elimination, our critical value is: **99.607** (see Table 5);
- Screening for outliners, since multiple regression is very sensitive regarding outliners;
- Making sure that we have linear relationship between the independent variables and the outcome.

Results

The result presented in this section are an initial indication of our research with this system; we are currently developing more detailed experiments and research with this rich dataset.

Experiment 1

Table 4 shows the results of the Kendall's tau coefficient, specifically the variant that makes adjustments for ties (*Tau-b*). Values of *Tau-b* range from -1 (100% negative association, or perfect inversion) to +1 (100% positive association, or perfect agreement). A value of zero indicates the absence of association.

Group	E	ES	A	C	O
All	−0.094	0.099	0.145	0.025	−0.379
Accepted	0.000	0.000	0.000	0.000	**0.800**
Rejected	−0.244	**0.333**	−0.067	0.022	−0.244
Reserved	0.010	0.010	0.162	0.153	−0.6

Table 4: Average rank correlation for applicant group versus personality trait (E: *Extraversion*; ES: *Emotional Stability*; A: *Agreeableness*; C: *Conscientiousness*; O: *Openness to Experience*)

The most significant positive relationship is between those applicants *Accepted* and the feature *Openness to Experience* (*Tau-b* = 0.8). A strong negative relationship exists

Class	Description	Potential Alias
A	Register early, and take some time to upload documents, but submit with plenty of time before deadline	EverythingEarly
B	Register reasonably early, but then upload documents and submit straight after with plenty of time before deadline, making no amendments	QuiteEarlyAndQuick
C	Similar to Class B, but submitting more slowly	Cautious
D	Registers reasonably early, and then takes time to upload, and only submits at the last days	VeryCautious
E	Latecomer to registration, but then uploads and submits quickly thereafter	Cautious
F	Latecomer to registration, but then uploads and submits slowly	Cautious
G	Latecomer to registration, but delays uploading and submission to last days	Cautious
H	Does everything at the last days, from registration to submission	EverythingLastMinute

Table 3: Description of each class

between those applicants *Reserved* and the feature *Openness to Experience* (*Tau-b* = -0.6).

Experiment 2

The following Figures 2–6 show box and whisper plots for each of the Five Factors, with the y-axis of each figure displaying the range for that particular feature. For example, Figure 2 displays the Extraversion feature, and the y-axis displays these values accordingly. The x-axis is comprised of the various classes from Figure 2, combined with the status of the application (1. *Accepted*, 2. *Rejected*, 3. *Reserved*, 4. *Ineligible*). Therefore, A1 are the Class A applicants who were *Accepted*, distinguished from A2, who were the same class (i.e. same activity based on timeline/milestones), but who were *Rejected*.

Experiment 3

We extracted the LIWC row data features (87 features) from the motivation letters and analysed the input data set against the 'status' of the application. The method used was multiple regression, the dependent variable being 'status' and the independent variables are the LIWC features (see Table 5).

UID	WC	WPS	UNIQUE	SIXLTR
1003	364	24.2667	41.4835	37.9121
1008	275	22.9167	61.4545	22.5455
1010	197	8.20833	68.0203	37.0558
1014	577	19.2333	53.7262	28.9428
1016	348	19.3333	55.4598	29.5977
1023	538	16.8125	53.9033	26.9517
1033	517	23.5	54.352	35.9768
1035	165	23.5714	62.4242	27.8788
1039	388	16.1667	56.1856	31.701
1040	491	14.8788	58.2485	33.4012
1049	462	25.6667	55.8442	33.1169
1058	293	32.5556	55.2901	26.9625
1069	436	29.0667	52.5229	26.1468
1073	162	27	61.1111	25.9259
1078	334	17.5789	55.988	34.4311

Table 5: Sample of the data set: we have 87 LIWC features and more than 1000 candidates. UID represent the user and rest of the columns represent the LIWC features

Result Set

The first result set shows the correlation between dependent variable (status) and independents variables (LIWC features). In our first assumption multicollinearity we use an R value of 0.7 or higher to say two predictable values have multicollinearity. Based on Table 5, if the tolerance is smaller than 0.1 then we have probability of multicollinearity, while VIF is the inverse of the tolerance, and so in the case of VIF greater than 10 then we have a case of multicollinearity. In this way, we found that the below LIWC features are related through multicollinearity:

```
REFERENCE_PEOPLE      SEEING
LEISURE_ACTIVITY      SELF
AFFECTIVE_PROCESS     WE
PHYSICAL_STATES       MUSIC
POSITIVE_EMOTION      TV_OR_MOVIE
SPORTS                FEELING
OTHER                 SLEEPING
BODY_STATES           HEARING
YOU                   SEXUALITY
SENSORY_PROCESS       OCCUPATION
HOME                  SOCIAL_PROCESS
PRONOUN               COGNITIVE_PROCESS
DIC                   NEGATIVE_EMOTION
```

Mahalanobis Distance

Checking Mahalanobis distance, we found that 102 records exceed the critical values, so we have removed these records since it is more than 2% of the total number.

Summary

Based on the result set we have, out initial data consisted of 1048 candidates and 86 LIWC features, before we build the multiple regression model we have gone through different checks to eliminate the features that may affect our model.

In the first assumption (Table 6), we removed 26 features because we detected its multicollinearity which will affect the model, and also eliminated 102 candidates with critical values in Mahalanobis distance check to reach the best model. On the model summary (Table 7), R Square shows that 6.9% accuracy to predicate the output of the "status" with these features. Signification (Table 8) of the model in ANOVA should be under 0.05 but the value is 0.187 which

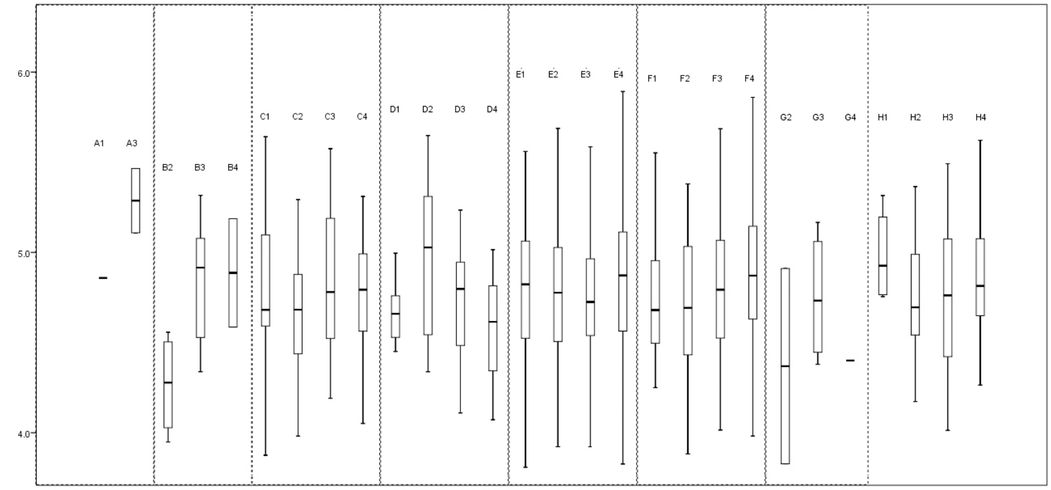

Figure 2: *Extraversion*. All features are hard to distinguish between, excepting that B2 is significantly smaller than B3 and B4

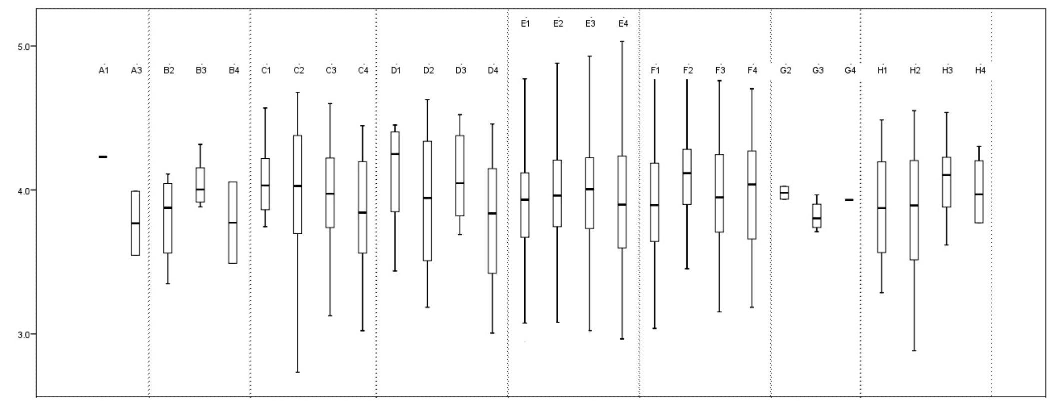

Figure 3: *Emotional Stability*. No real features larger or smaller, although the range on all of the E features seems much greater than the other features

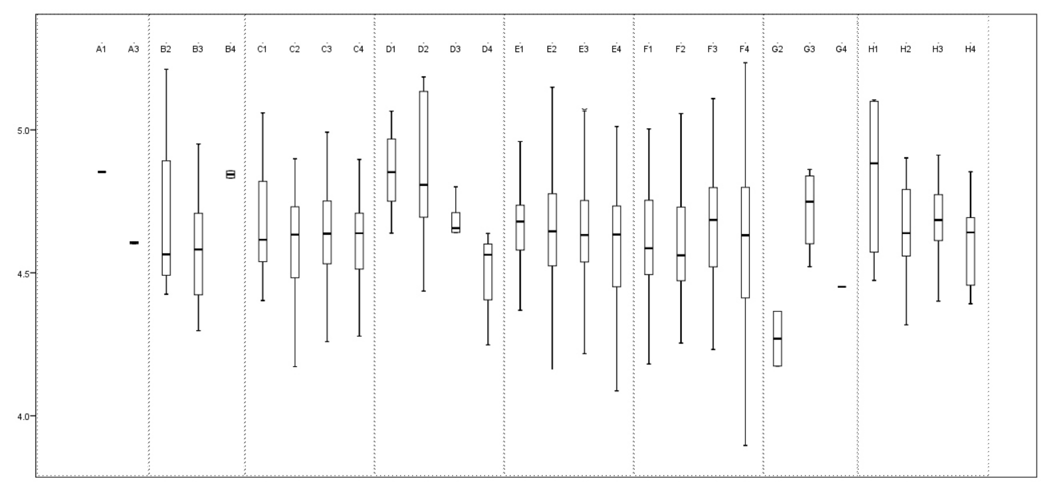

Figure 4: *Agreeableness*. D4 is significantly smaller than D1, D2, and D3. G2 appears significantly less conscientious than G3

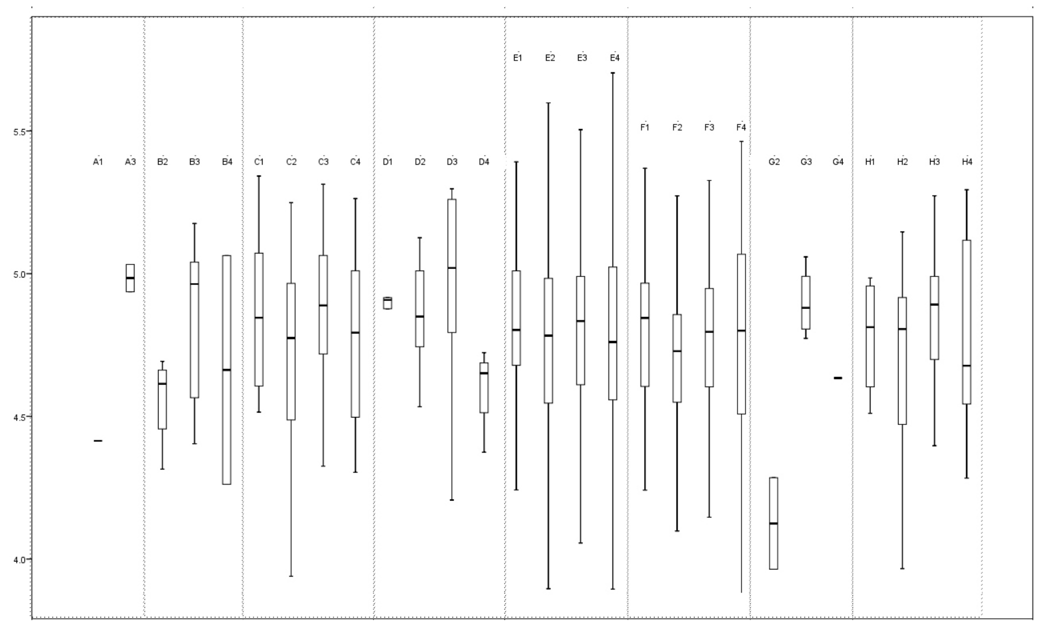

Figure 5: *Conscientiousness*. G2 appears significantly less conscientious than G3. To a lesser degree D4 is smaller than D1, D2, and D3

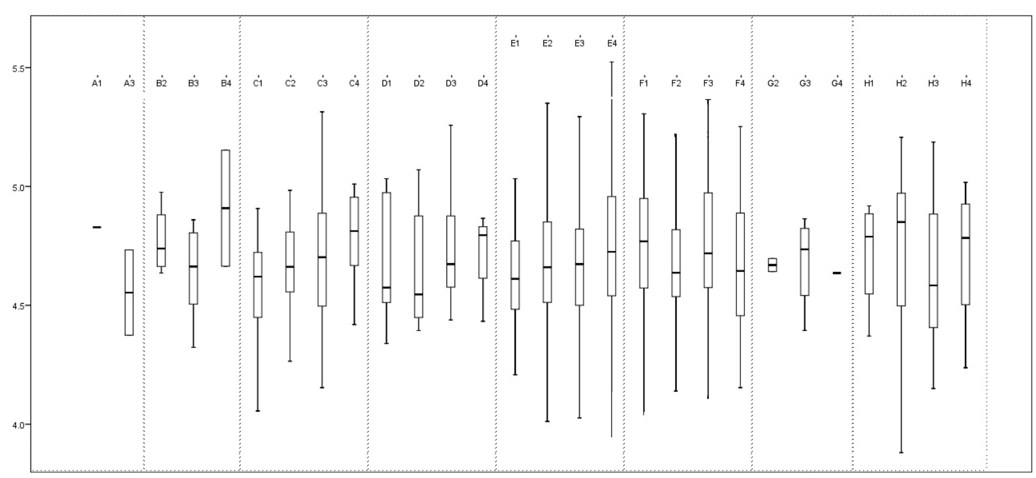

Figure 6: *Openness to Experience*. As with Emotional Stability, there are no exceptional features, although the range on all of the E features seems much greater than the other features. The Class E were the applicants that were relative late comers to registration, but who then uploaded and submitted quickly thereafter. Openness to Experience would seem to have very little relationship with this class of applicant

Model		*Collinearity Statistics* Tolerance	VIF
1	(Constant)	-	-
	REFERENCE_PEOPLE	0.003	290.011
	LEISURE_ACTIVITY	0.004	266.927
	AFFECTIVE_PROCESS	0.006	181.557
	PHYSICAL_STATES	0.006	181.512
	POSITIVE_EMOTION	0.006	176.569
	SPORTS	0.007	139.147
	OTHER	0.007	138.914
	BODY_STATES	0.008	122.386
	YOU	0.009	108.779
	SENSORY_PROCESS	0.011	89.824
	HOME	0.015	67.555
	PRONOUN	0.022	46.061
	DIC	0.023	43.246
	SEEING	0.025	39.690
	SELF	0.029	34.103
	WE	0.033	30.494
	MUSIC	0.037	27.254
	TV_OR_MOVIE	0.041	24.260
	FEELING	0.043	23.210
	SLEEPING	0.044	22.841
	HEARING	0.047	21.488
	SEXUALITY	0.049	20.536
	OCCUPATION	0.050	19.857
	SOCIAL_PROCESS	0.070	14.252
	COGNITIVE_PROCESS	0.075	13.407
	NEGATIVE_EMOTION	0.089	11.213

Table 6: Coefficients of multicollinearity variance influence factor

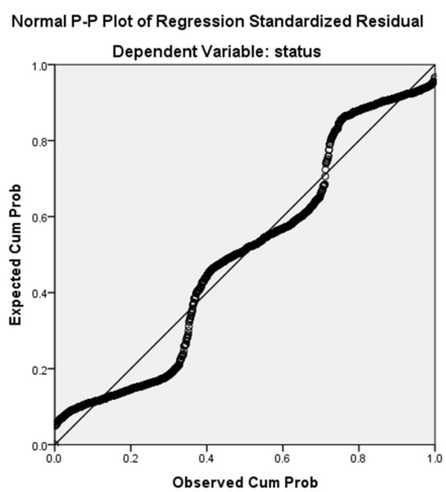

Figure 7: Normal P-P Plot

means that the current model cannot produce accurate prediction for the output. The most effective feature in the model in Table 9 are NEGATIONS.

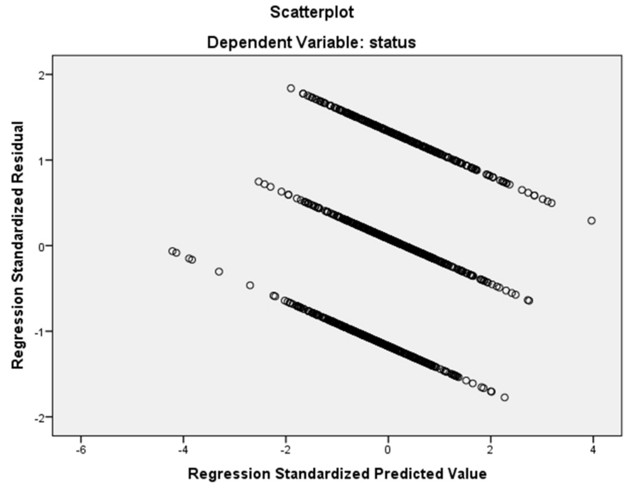

Figure 8: Scatterplot

Model	R	R Sq.	Adj. R Sq.	Std. Err of Est.
1	.262[a]	0.069	0.010	0.796

Table 7: Model summary after removing the multicollinearity features and above critical value of Mahalanobis distance

	Sum of Sq.	df	Mean Sq.	F	Sig.
Regress.	41.602	56	0.743	1.172	0.187[b]
Residual	563.465	889	0.634	-	-
Total	605.067	945	-	-	-

Table 8: Evaluation of the model and ability to predicate the status values

Model		Stnd. Coeff. Beta	Sig.
1	(Constant)		0.000
	NEGATIONS	0.109	0.004
	QMARK	0.107	0.199
	SPACE	0.098	0.044
	ABBREVIATIONS	0.076	0.038
	CAUSATION	0.073	0.051
	DASH	0.068	0.093
	UNIQUE	0.068	0.247
	INHIBITION	0.062	0.072
	JOB_OR_WORK	0.051	0.188
	DISCREPANCY	0.045	0.289
	SCHOOL	0.044	0.279

Table 9: Top effective coefficient LIWC features over the model

Conclusions and Future Work

We have started the analysis of interactions with our chosen information system utilising a small set of features; there are many more that we are planning to investigate. We list below the features we have used in these experiments, and the remaining features that can be used in future experiments:

- **Data used in these experiments:** Five Factors on text from Facebook Interaction; Five Factors on text from motivation letters; Status (e.g. Eligibility, Final Selections); Timeline/Footprint.
- **Data available but not used:** Complete/detailed Timeline/Footprint; Complete/detailed Facebook profile which includes interactions between applicants; Demographic Information (e.g. gender, age); Appeals after receiving application decision (e.g. email or webpage help text); Five Factors Questionnaire (for current calls); Course Details; Course Favourites; Email interactions during application; Evaluation Grades (actual scores, broken down into categories); Additional forms (qualifications, CV, research proposal); Reasons for ineligibility (e.g. documents missing, already received scholarship within 12 months); Mobility Levels (e.g. undergraduate, postgraduate); Mobility Type (e.g. exchange, degree seeking); Host Institutions (where they applied).

The results from our two streams of experiments or explorations provide a valuable initial indication about the data and an appropriate way to represent and explore it. From Experiment 1, there seems to be a strong relationship between the Five Factor feature *Openness to Experience* with a strong correlation with the *Accepted* group. The exploration of timeline behaviour is dependent on our representation used for interactions, and the classes derived. This is a first attempt at tackling this representation problem. The same feature *Openness to Experience* has no group/class combinations that are significantly different than others.

The investigation using the Five Factors is very much in its initial stages, and we wish to develop further to deliver an appropriate instrument to the applicants to determine precisely their trait values, and then we can explore how different personality types interact. For this initial study we have utilised the only method available to us, which is that determined by Mairesse et al. (2007). Our future work will continue to focus in developing the multiple regression model based upon more accurate data by including more details about the user instead of only the motivation letter.

We are continuing with our dictionary-based approach to examine content differences (LIWC), psycholinguistic differences (MRC), and will also include affect differences using the Dictionary of Affect in Language (DAL) (Whissell 2008; 2009). The DAL is designed to measure the emotional meaning of words and texts, comparing against a word list rated by people for their activation, evaluation, and imagery. Additionally we are using the SentiWordNet (Esuli and Sebastiani 2006) and WordNet-Affect (Strapparava and Valitutti 2004) extensions to WordNet (Miller 1995). WordNet is a large lexical database of English. Nouns, verbs, adjectives and adverbs are grouped into sets of cognitive synonyms (synsets), each expressing a distinct concept. Synsets are interlinked by means of conceptual-semantic and lexical relations. SentiWordNet is a lexical resource for opinion mining. WordNet-Affect labels synsets with a hierarchical set of emotional categories.

We will correlate these with the Five Factors and trait emotional intelligence (Petrides and Furnham 2001) through administration of the Big Five Inventory (BFI, 44-item) (John, Donahue, and Kentle 1991) and Trait Emotional Intelligence Questionnaire–Short Form (TEIQue-SF, 30-item) questionnaires designed to measure global trait emotional intelligence (Petrides and Furnham 2006) respectively. In this way we determine our own domain specific formulae for traits and behaviours, instead of relying on the bias inherent in those from Mairesse et al. (2007). In particular, we are especially interested in the temporal aspects of traits and affects, which we plan to explore using Allen's logic of time and action (Allen 1983; 1984).

References

Allen, J. F. 1983. Maintaining knowledge about temporal intervals. *Communications of the ACM* 26(11):832–843.

Allen, J. F. 1984. Towards a general theory of action and time. *Artificial Intelligence* 23(2):123–154.

Blamey, B.; Crick, T.; and Oatley, G. 2012. R U :-) or :-(? Character- vs. Word-Gram Feature Selection for Sentiment Classification of OSN Corpora. In *Research and Development in Intelligent Systems XXIX, Proceedings of the 32nd SGAI International Conference on Artificial Intelligence (AI-2012)*, 207–212. Springer.

Blamey, B.; Crick, T.; and Oatley, G. 2013. 'The First Day of Summer': Parsing Temporal Expressions with Distributed Semantics. In *Research and Development in Intelligent Systems XXX, Proceedings of the 33rd SGAI International Conference on Artificial Intelligence (AI-2013)*, 389–402.

Esuli, A., and Sebastiani, F. 2006. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-06)*, 417–422.

Goldberg, L. 1990. An alternative "description of personality": The big Five Factor structure. *Journal of Personality and Social Psychology* 59(6):1216–1229.

Iacobelli, F.; Gill, A. J.; Nowson, S.; and Oberlander, J. 2011. Large Scale Personality Classification of Bloggers. In *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction (ACII 2011)*, volume 6975 of *Lecture Notes in Computer Science*, 568–577. Springer.

John, O.; Donahue, E.; and Kentle, R. 1991. The Big Five Inventory–Versions 4a and 54. Technical report, University of California, Berkeley, Institute of Personality and Social Research.

Kendall, M. 1938. A New Measure of Rank Correlation. *Biometrika* 30(1-2):68–80.

Mairesse, F.; Walker, M.; Mehi, M.; and Moore, R. 2007. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *Journal of Artificial Intelligence Research* 30:457–500.

Miller, G. 1995. WordNet: A Lexical Database for English. *Communications of the ACM* 38(11):39–41.

Norman, W. T. 1963. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology* 66(6):574–583.

Oatley, G., and Crick, T. 2014. Changing Faces: Identifying Complex Behavioural Profiles. In *Proceedings of 2nd International Conference on Human Aspects of Information Security, Privacy and Trust (HAS 2014)*, volume 8533 of *Lecture Notes in Computer Science*, 282–293. Springer.

Peabody, D., and Goldberg, L. 1989. Some determinants of factor structures from personality-trait descriptor. *Journal of Personality and Social Psychology* 57(3):552–567.

Pennebaker, J.; Francis, M.; and Booth, R. 2001. Linguistic Inquiry and Word Count. Erlbaum Publishers.

Petrides, K., and Furnham, A. 2001. Trait emotional intelligence: psychometric investigation with reference to established trait taxonomies. *European Journal of Personality* 15(6):425–448.

Petrides, K., and Furnham, A. 2006. The role of trait emotional intelligence in a gender-specific model of organizational variables. *Journal of Applied Social Psychology* 36(2):552–569.

Strapparava, C., and Valitutti, A. 2004. WordNet-Affect: an Affective Extension of WordNet. In *Proceedings of the 4th Conference on Language Resources and Evaluation (LREC-04)*, 1083–1086.

Vazire, S., and Gosling, S. 2004. e-Perceptions: Personality Impressions Based on Personal Websites. *Journal of Personality and Social Psychology* 87(1):123–132.

Whissell, C. 2008. A comparison of two lists providing emotional norms for English words (ANEW and the DAL). *Psychological Reports* 102(2):597–600.

Whissell, C. 2009. Using the Revised Dictionary of Affect in Language to quantify the emotional undertones of samples of natural language. *Psychological Reports* 105(2):509–521.

Wilson, M. 1988. The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2.00. *Behavior Research Methods, Instruments & Computers* 20(1):6–10.