UNIVERSITY OF LEEDS

This is a repository copy of *Diversity Profiling of Learners to Understand Their Domain Coverage While Watching Videos*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/134064/

Version: Accepted Version

Proceedings Paper:
Abolkasim, E, Lau, L orcid.org/0000-0003-1062-9059, Dimitrova, V et al. (1 more author) (2018) Diversity Profiling of Learners to Understand Their Domain Coverage While Watching Videos. In: Lifelong Technology-Enhanced Learning. EC-TEL 2018: 13th European Conference on Technology Enhanced Learning, 03-06 Sep 2018, Leeds, UK. Lecture Notes in Computer Science, 11082 . Springer , pp. 561-565. ISBN 978-3-319-98571-8

https://doi.org/10.1007/978-3-319-98572-5_45

© 2018, Springer Nature Switzerland AG . This is a post-peer-review, pre-copyedit version of an article published in Lifelong Technology-Enhanced Learning. The final authenticated version is available online at: https://doi.org/10.1007/978-3-319-98572-5_45. Uploaded in accordance with the publisher's self-archiving policy.

Reuse
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose
university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Diversity Profiling of Learners to Understand Their Domain Coverage While Watching Videos

Entisar Abolkasim[1], Lydia Lau[1], Vania Dimitrova[1] and Antonija Mitrovic[2]

[1] School of Computing, University of Leeds, UK
[2] Intelligent Computer Tutoring Group, University of Canterbury, New Zealand

Abstract. Modelling diversity is especially valuable in soft skills learning, where contextual awareness and understanding of different perspectives are crucial. This paper presents an application of a diversity analytics pipeline to generate domain diversity profiles for learners as captured in their comments while watching videos for learning a soft skill. The datasets for analysis were collected from a series of studies on learning presentation skills with Active Video Watching system (AVW-Space). Two user studies (with 37 postgraduates and 140 undergraduates, respectively) were compared. The learners' diversity and personal profiles are used to further understand and highlight any notable patterns about their domain coverage on presentation skills.

Keywords: diversity profiling, domain coverage, diversity analytics pipeline, video-based learning, presentation skills

1 Introduction

Videos are widely used by learners and tutors as a prime medium for learning and teaching [4]. Videos can be especially powerful for soft skills training. If carefully chosen, it can provide opportunities for self-regulated learning and for exploring different perspectives on the same situation. The paper presents a novel computational approach to automatically detect the domain coverage in learner comments by deriving diversity profiles for learners, and investigates how this may relate to individual learner's characteristics. The Semantic-Driven Diversity Analytics Tool (SeDDAT) presented in [1] has been extended and applied on new datasets, obtained from two user studies with postgraduate and undergraduate students respectively, on learning presentation skills from videos.

By adapting the Stirling Diversity Framework [6], domain diversity profiles for the learners are generated in terms of variety, balance and disparity. Variety refers to breadth of domain coverage. This is useful for learning to gather the learners' overview of the domain. Balance goes further and captures the extent of domain coverage. This is useful to see the degree of consistency in the level of understanding across domain categories. Disparity refers to the density of domain coverage, i.e. measures how spread out the domain concepts are covered by the pool of comments. The learners' diversity and personal profiles are then analysed to address the following research question:

Are there any notable differences between user groups with regards to domain diversity and individual profiles?

2 Overview of the Diversity Profiling Pipeline

The pipeline consists of: **a)** Input preparation for SeDDAT- including the appropriate ontology, an entry point and content file(s) annotated using this ontology. **b)** Diversity measurements using SeDDAT to create diversity profiles. **c)** Diversity Analysis, where the analyst inspects the diversity profiles for diversity patterns. **d)** Fine-tuning of profiling, if interesting patterns are spotted, a new entry point for SeDDAT can be specified for further diversity analysis. More details are described in [2].

3 Datasets for Profiling

The diversity pipeline was applied on two user studies using the Active Video Watching (AVW-Space). **Study A** had 37 postgraduate students (**PGs**) and **Study B** 140 undergraduate students (**UGs**); details about the design of studies can be found in [3]. The following collected data were specifically relevant to this research: **(i)** data about the videos; **(ii)** user profiles, such as demographics, background experiences, Motivated Strategies for Learning Questionnaire (MSLQ) [5]; and **(iii)** user comments. The total number of comments was 744 from Study A and 1129 from Study B.

A Presentation Skills Ontology (PreSOn) was developed (as described in [2]) to automatically tag the user comments and assist diversity profiling. The semantic tagging resulted in a total of 1,217 annotations for Study A and 2,070 for Study B; with 197 and 220 distinct entities, respectively. The average number of annotations and distinct entities per video covered by the comments are: Study A - 152.1 (std. 30.1) and 66 (std. 8.7); and Study B - 258.8 (std. 65.0) and 78.5 (std. 9.1).

4 Domain Diversity Profiling for Learners

The user (learner) diversity profiles and other associated data (e.g. demographics, MSLQ, knowledge, etc.) are analysed below to address the research question.

Comparing the study groups. Learners' personal and diversity profiles were compared to see if PGs differ from UGs in their background knowledge, attitudes towards learning and their behavior during video watching. Table 1 reports the profile items with significant difference between the two studies.

PGs seem to have higher domain diversity (variety, balance and disparity) compared to UGs – i.e. PGs had more diverse domain coverage with regards to presentation skills. PGs reported significantly more training and experience on presentation skills. This may explain the higher diversity scores. PGs scored higher on MSLQ Task Value, Intrinsic Motivation, Organisation, Elaboration and Self-Regulation, whereas UGs scored higher on Extrinsic Motivation, Academic Control and Effort regulation. These figures seem to correlate to the fact that Study A was on a voluntary basis (more comments) whereas Study B was part of a course assessment (fewer comments).

Comparing personal attributes. Each diversity property was compared across all learners' personal attributes to see if the learners' personal characteristics, such as gender or language (native/non-native speaker), will influence diversity scores. The only significant difference was in gender for Study A. The domain balance ($U = 208$, $p < .05$) and domain disparity ($U = 205$, $p < .05$) were significantly higher for female learners ($n = 26$) than male learners ($n = 11$). It is surprising that the language attribute did not have an impact on the coverage of the domain (i.e. diversity scores).

Table 1. Significant differences between learners from Studies A and B; † denotes Likert scale was used - 1 (lowest) to 5 (highest); ** significance at $p < .005$, and * at $p < .05$.

User profile items	Study A (37)	Study B (140)	Significance
Domain Variety	3.65 (.79)	2.91 (1.13)	t = 4.55**
Domain Balance	.49 (.24)	.28 (.22)	t = 4.98**
Domain Disparity	10.23 (4.43)	7.88 (5.4)	t = 2.73*
Comments (no. of)	19 (12.87)	7.89 (9.59)	t = 4.91**
Training†	2.14 (.95)	1.7 (.78)	t = 2.57*
Experience†	2.86 (.79)	2.34 (.85)	t = 3.53**
Task Value†	4.50 (.38)	3.97 (.59)	t = 6.69**
Academic Control†	3.91 (.47)	4.14 (.57)	t = 2.64*
Intrinsic†	4.05 (.53)	3.77 (.59)	t = 2.82*
Extrinsic†	3.41 (.81)	4.19 (.63)	t = 5.48**
Effort Regulation†	2.95 (.43)	3.56 (.66)	t = 6.84**
Organisation†	3.85 (.95)	3.22 (.89)	t = 3.63**
Elaboration†	4.13 (.55)	3.67 (.66)	t = 4.33**
Self-Regulation†	3.61 (.39)	3.28 (.52)	t = 4.19**

Comparing the most and least diverse learners. To further understand if any of the learners' profile items contribute to the domain coverage, their diversity scores are ranked for each study - variety, then balance, and then disparity. The top and bottom quartile were analysed versus the middle two quartiles. There are significant differences on the number of comments in both studies (Table 2) between these three subgroups of learners (all pairwise comparisons significant at $p < .005$). In Study A, there was also significant difference between the subgroups based on presentation experience. Although it was expected that experienced learners should have higher diversity properties, this was not the case.

Table 2. Comparing quartiles defined on domain variety, balance and disparity (all pairwise comparisons significant at $p < .005$).

	Top quartile Study A (9) Study B (35)	Middle quartiles Study A (19) Study B (70)	Bottom quartile Study A (9) Study B (35)	Significance
Experience Study A	2.56 (1.13)	3.16 (.5)	2.56 (.73)	W = 7.664*
Comments Study A	34.33 (13.19)	17 (6.77)	7.89 (7.42)	W = 20.64**
Comments Study B	18.06 (13.94)	5.83 (3.61)	1.83 (1.36)	W=83.46**

Comparing correlations. For both studies, domain variety, balance and disparity are strongly correlated (with correlations ranging from .494 to .904, all at $p<.005$). Also, the number of comments is strongly correlated with domain variety, balance and disparity in both studies ($p < .005$). This indicates that learners should be triggered to write more comments, as the more they write the more they notice with regards to the domain while watching the videos.

8. Conclusion

This paper presented an instantiation of a novel semantic driven analytics pipeline for understanding domain diversity in learners' comments when watching videos. SeD-DAT was applied on two studies about presentation skills to generate domain diversity profiles. PGs had significantly higher domain diversity than UGs; the former were more intrinsically motivated while the latter were more extrinsically motivated. It was surprising that the native language did not impact on the domain diversity, which indicated that cognitive understanding of presentation skills was orthogonal to language. This work contributes to future intelligent learning environments that address the needs of the learners and their diverse background in the modern society, which would require automated ways to capture and compare different domain perspectives.

Acknowledgements. Support by EU-FP7-ICT-257184 ImREAL, Ako Aotearoa and teaching development grant at University of Canterbury. We thank Amali Weerasinghe and Alicja Piotrkowicz for helping with the creation of PreSOn, and the expert trainers from Skills@Library at Leeds University for validating the ontology. Finally, we thank all participants in the user studies.

References

1. Abolkasim, E., Lau, L., & Dimitrova, V. (2016). A Semantic-Driven Model for Ranking Digital Learning Objects Based on Diversity in the User Comments. In European Conference on Technology Enhanced Learning, Springer, Cham, 3-15.
2. Abolkasim, E., Lau, L., Dimitrova, V., & Mitrovic, A. (in press). Ontology-based Domain Diversity Profiling of User Comments. In Proceedings of the 17th Conference on Artificail Intelligence in Education.\
3. Dimitrova, V., Mitrovic, A., Piotrkowicz, A., Lau, L., & Weerasinghe, A. (2017). Using learning analytics to devise interactive personalised nudges for active video watching. In Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization . ACM, 22-31.
4. Hua, K. (2015). Education as Entertainment: YouTube Sensations Teaching The Future. Forbes.
5. Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. Journal of educational psychology, 82(1), 33.
6. Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. Journal of the Royal Society Interface, 4(15), 707-719.