Enhancing Digital Literacy by Multi-modal Data Mining of the Digital Lifespan

John Collomosse
Stuart James
CVSSP
University of Surrey
Guildford, UK
j.collomosse@surrey.ac.uk

Abigail Durrant
Diego Trujillo-Pisanty
Culture Lab
Newcastle University
Newcastle upon Tyne, UK
abigail.durrant@ncl.ac.uk

Wendy Moncur
Kathryn M. Orzech
University of Dundee
Dundee, UK
wmoncur@dundee.ac.uk

Sarah Martindale
Horizon
University of Nottingham
Nottingham, UK
Sarah.Martindale@nott.ac.uk

Mike Chantler
Heriot-Watt University
Edinburgh, UK
m.j.chantler@hw.ac.uk

ABSTRACT

Social media pervades modern digital society, yet despite its regular use the general level of digital literacy and awareness of online representations of self remain low. We report progress on the RCUK DE funded 'Charting the Digital Lifespan' project, which promotes digital literacy of social media through design of novel technological interventions that raise awareness of the digital personhood. We describe progress towards automatic social media data-mining capable of categorising photographs and associated comments into high-level semantic concept groups elicited from digital anthropological study. We describe how this pilot system has been incorporated into live trials of technologies that visualize and encourage reflection upon digital personhood.

Categories and Subject Descriptors

D.4 [**Human-centered computing**]: Social Media

1. INTRODUCTION

Over two billion individuals world-wide, and 57% of the UK population, maintain digital reflections of their analogue lives through the lens of social media [1]. The paths of these digital and physical lives run in parallel, converging and diverging as they mediate personhood. The RCUK Digital Economy Theme funded Charting the Digital Lifespan (CDL) project is embarking upon the second half of a two year interdisciplinary study exploring this space. A key goal for CDL is to enhance the digital literacy of individuals by facilitating reflection upon their digital personhood. Frequently individuals are unaware of the permanence of their digital footprints online, which are often contributed in an ad-hoc incremental manner without regard to the holistic representations being constructed of their digital selves. Repercussions can occur when these digital representations are revealed to, or interact with, others in unexpected ways.

1.1 Methodology and Context

CDL studies drivers and use patterns of social media at three life stages: emerging adults; new parents; and retirees. An experience-centred design (ECD) [13] methodology is adopted, in which technology probes [8] — hardware or software devices — are introduced into individuals' lives enabling novel forms of reflection upon their digital personhood. The probes provoke individual reflection by presenting a visualisation of the digital personhood, instantiated through automatic analysis ('data mining') of online social media presence. The design of the technology probe itself is informed by both semi-structured interviews conducted with individuals also at that life-stage (digital anthropology [7]), and exploratory workshops centred around *design fictions* that encourage critical speculation about future possibilities involving digital personhood [2]. The needs of the technology probe drive technical innovation in social media data mining, which forms the focus of this paper.

One concrete example of the CDL methodology was the 'Admixed Portrait' [12]; a visualisation aimed at promoting reflection on online identity in the new parent user-group. For this technology probe, a visual amalgam was generated by data mining face portrait images recently posted by a participant to their Facebook timeline (Fig. 1). New parents were able to reflect upon how they are portrayed online, through the nature and content of their posts (e. g. a trend towards posts of their new child) and posts made by others. This intervention was designed around new parent use cases identified through anthropological study.

In our current instantiation with the emerging adult group, we are encouraging reflection on the kinds of material previously shared on their Facebook newsfeed. An exploratory user study was used to determine a set of nine representative 'concept groups', e. g. sport, friends and family (Sec. 3.1). Computer Vision (CV) and Machine Learning (ML) algorithms have been developed to automatically classify participants' social media posts into one or more of these groups. Due to the challenges automatic classification, only posts of photographs (visuals) with associated free text comments were considered. Visualizations indicating the frequency with which posts were made into each group were then shown to participants to encourage reflection on the kinds of activity that they were depicting through their digital personhood. We elaborate upon the CV/ML algorithms employed in performing the post classification, and the results obtained, in the remainder of this paper.

2. RELATED WORK

The identification of semantic concepts within images (*image classification*) is a fundamental CV/ML challenge. Modern approaches tackle the problem using a three-stage pipeline. First, a set of features are extracted from the image encoding texture and visual structure in a high dimensional space. Commonly features are gradient domain (e. g. SIFT [9], HOG [5]). Second, the feature space is simplified whilst maintaining its ability to discriminate content. This is typically performed by clustering using k-means [11] or Gaussian mixtures [10] to form a visual dictionary, and expressing all features with respect to that dictionary. Third, the image descriptors for each concept are used to train a classifier using a set of training examples. More recently the pipeline has been adapted to learn the classifier and appropriate first-stage features simultaneously; so called 'deep learning' [4]. These frameworks have been applied classically to lab-scale datasets; a few tens of videos, or thousands of images with success rates approaching 70% for object recognition e. g. PASCAL VOC [10]. More recently large datasets such as ImageNet [6], containing many hundreds of categories each with tens of exemplars have been explored. In our setup we face the complementary problem of only a *few* categories but *many thousands* of diverse exemplars (high intra-class variance) and sparse text data that previous, domain constrained, social classification work does not accommodate.

3. SOCIAL MEDIA CLASSIFICATION

We adopt a standard *supervised* classification approach in which a set of social media posts (SMPs) are marked up manually by tagging any of the nine concepts present. A fraction of this data is used to train our proposed classifier, and the remainder used to test the system to evaluate accuracy. The train-test split is randomised and repeated several times in a cross-validation framework to report an mean averaged precision (MAP) value for accuracy.

3.1 Datasets and Concept Groups

Two datasets were harvested from Facebook using a bespoke web crawler. The first indexed private Facebook profiles of ~ 20 participants in the emerging adult group. The second indexed publicly available profiles linked from emerging adults. Approximately ~ 47k posts were harvested, and ~ 6k usable records containing both photos and English text comments were retained for the combined dataset. Manual annotation was performed to establish ground-truth. For public data, the CDL team annotated the data. For private data, annotation was crowd-sourced from participants contributing their data.

The nine concept groups for annotation and subsequent classification were scoped initially by CDL staff, and subsequently refined during 15 semi-structured interviews with emerging adults conducted by the project anthropologist. Participants volunteered category names, and these were grouped after all interviews with emerging adults were complete, aligning these with the preliminary concept groups when appropriate. The resulting concept groups were: Art; Attitude & Beliefs; Family & Pets; Food; Friends; Travel; Celebrations; Personal style and self-imagery (e. g. selfie); Sports. To collect private Facebook data and have it marked up, the project anthropologist enlisted first-year art students to submit their Facebook image data and classify images donated by their peers to explore images categories. This activity took place in the context of a design project aimed at creating a novel way to share digital photos. The students had the option to opt-out of the data collection and classification, but most (22) participated, allowing a Facebook app designed within the project to access their photos and associated comments. A web interface presented random photos to participants, who chose appropriate category labels.

3.2 Text mining

For each SMP a set of nouns, verbs and adjectives were extracted from text comments provided by both the author and social media contributors. Topic discovery was performed using LDA based on keyword co-occurrence [3], using the set of keywords present for each annotated concept group. The result was a set of 10 representative keywords for each concept group, determined by SMP content. A binary feature space was defined over the resulting 9×10 dimensional set of keywords, with each dimension indicating the presence of a word in SMP comments. Non-linear support vector machines were trained in a one-vs-all pattern using SMPs from the training class. An mean average precision (MAP) of 32.0% was obtained (Fig. 2, green).

3.3 Visual mining

Dense color-SIFT features were extracted from each photograph in the SMP training set, and 10% of these were clustered using $k-$means to quantize the feature space in to $k = 2000$ codewords. Histograms of codeword occurrence were built for each training SMP yielding a Bag of Visual Words [11] representation for the media, which were used to train a non-linear support vector machine in a one-vs-all pattern. A MAP of 25.3% was obtained (Fig. 2, blue).

3.4 Multimodal fusion

An iterative fusion method adapted from Zhang et al.'s Co-Trade [14] was used to integrate both the trained Text and Visual SVM classifiers to produce a combined classifier informed by both modalities. Training SMPs were partitioned further, into a Training and Validation set on a 50:50 basis. Validation set SMPs were labelled using the resulting classifiers. A hyper-graph was built for each modality connecting each training and validation SMP (node) with its K closest neighbours (in our experiments $K = 5$), as measured by Euclidean distance in the respective modality. Each node was evaluated for label coherency with its immediate neighbours, and ranked. An initially empty set of 'coherent' training data was incrementally built by steadily adding the most coherent nodes from the unified training and validation set, until classification accuracy over validation data reached a maximum. This resulted in a single, overall classifier with performance comparable (a fraction of a percentage point greater) than the better of the two uni-modal classifiers. The reflected the complementarity of performance seen by those independent classifiers (Fig. 2).

4. DISCUSSION AND CONCLUSIONS

Social media is highly diverse, complicating efforts to automatically categorise it. A system has been outlined to attempt this through multi-modal fusion of visual and text classifiers, yielding performance of $\sim 30\%$ MAP on unstructured raw social media in the wild. Whilst encouraging, the system does not yet generalise to accommodate unseen data well – e. g. new keywords ('Whiskey' would not be regarded as similar to the word 'drink', in the celebrations concept). We are augmenting our system to incorporate WordNet as a semantic distance measure in our text classifier to enable such inferences. Visual classification is also a challenge, due to noise in the training set and difficulty of obtaining a sufficiently large quantity of training mark-up to counter the high visual diversity. We are exploring domain adaptation from web-based sources, and user-specific annotation (e. g. in the form of user supplied 'corrections' to categorisations) to improve accuracy. The application of our classification is visualizations to facilitate reflection on the digital lifespan. As such, 'perfect' classification is not needed – rather, a 'mostly correct' result is acceptable to visualise general trends in SMPs which tend to be the most effective form of data presentation within the technology probes we continue to explore within CDL.

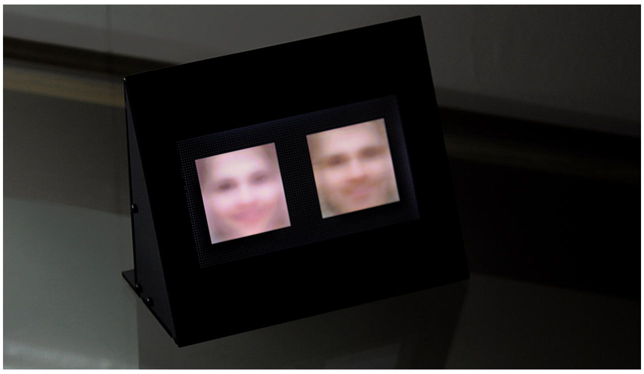

Figure 1: The Ad-mixed Portrait, used to promote reflection amongst new parents on their digital representations of self on social media (Facebook) [12].

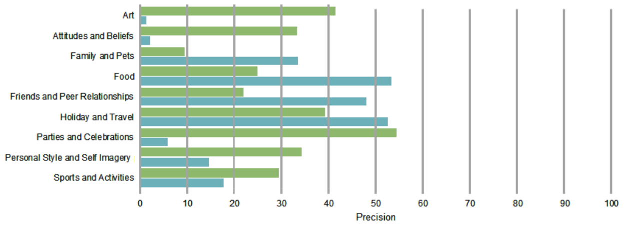

Figure 2: Result of independent modalities used to classify SMPs in the wild: Text (green) and Visual (blue) yielded independent accuracies of 32.0% and 25.3% MAP respectively, and exhibited complementary performance across the categories. This indicates value in pursuing a multi-modal SMP classification strategy.

Acknowledgements

The Charting the Digital Lifespan (CDL) project is funded by RCUK (EP/L00383X/1). The third author is additionally supported by The Leverhulme Trust (ECF-2012-642).

5. REFERENCES

[1] Internet access - households and individuals 2012. Technical report, Office of National Statistics, London, UK, February 2013.

[2] J. Bleecker. Design fiction: A short essay on design, science, fact and fiction. `http://nearfuturelaboratory.com/2009/03/17/design-fiction-a-short-essay-on-design-science-fact-an` Accessed: 2014-09-12.

[3] D. Blei, A. Ng, and M. Jordan. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 2003.

[4] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. Return of the devil in the details: Delving deep into convolutional nets. In *Proceedings of the British Machine Vision Conference (BMVC)*, 2014.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proceedings Intl. Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 886–893, 2005.

[6] T. Deselaers and V. Ferrari. Visual and semantic similarity in imagenet. In *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1777–1784, 2011.

[7] H. A. Horst and D. Miller. *Digital Anthropology.* Berg, London, 2012.

[8] H. Hutchinson, W. Mackay, B. Westerlund, A. Druin, C. Plaisant, M. Beaudouin-Lafon, S. Conversy, H. Evans, H. Hansen, N. Roussel, and B. Eiderback. Technology probes: inspiring design for and with families. In *Proceedings of ACM CHI*, pages 17–24, 2003.

[9] A. Lowe. Object recognition from local scale-invariant features. In *Proceedings Intl. Conf. on Computer Vision (ICCV)*, pages 1150–1157, 1999.

[10] J. Sanchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the fisher vector: Theory and practice. Technical Report 8209, INRIA, Project LEAR, Grenoble, France, May 2013.

[11] J. Sivic and A. Zisserman. Video Google: Efficient visual search of videos. In J. Ponce, M. Hebert, C. Schmid, and A. Zisserman, editors, *Toward Category-Level Object Recognition*, volume 4170 of *LNCS*, pages 127–144. Springer, 2006.

[12] D. Trujillo-Pisanty, A. Durrant, S. Martindale, S. James, and J. Collomosse. Admixed portrait: Reflections on being online as a new parent. In *Proceedings of ACM Designing Interactive Systems (DIS)*, pages 503–512. ACM Press, 2014.

[13] P. Wright and J. McCarthy. *Experience-Centred Design: Designers, Users and Communities in Dialogue.* Morgan & Claypool, 2010.

[14] M. Zhang and Z. Zhou. Cotrade: Confident co-training with data editing. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 41(6), December 2011.