Play with Me - Measuring a Child's Engagement in a Social Interaction

Shyam Sundar Rajagopalan[1], O.V. Ramana Murthy[1], Roland Goecke[1], Agata Rozga[2]

FG 2015 – 6 May 2015

[1]Vision & Sensing Group, Human-Centred Computing Lab, Faculty of ESTeM, University of Canberra, Australia

[2]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA

UNIVERSITY OF CANBERRA

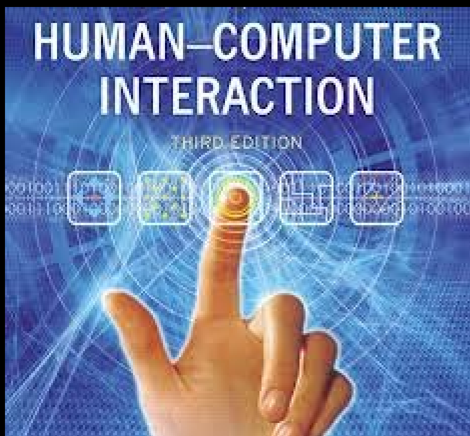
Computational Behaviour Analysis in Social Interactions - Challenges

1. Human behaviours expressed using subtle verbal and non-verbal cues are often contextual and social.

2. Analysis of interaction dynamics using multimodal signals.

3. Hard to obtain non-invasive high-level features such as body and head poses in semi-structured social interactions.

Applications



...

Source : Google Images

Autism Spectrum Disorders (ASD)

What
social, communication and cognitive deficits

Statistics
Fastest growing
1 in 88

Treatment
Early Intervention

Diagnosis
No single genetic/ biological marker
Behaviour study
Detection > 18-24 months

Source : CDC and Autism Speaks

Red flags by 14 months
Low social responsiveness and reciprocity
Infrequent initiation of joint attention and response to social cues
Infrequent playful imitation
Language delay
Repetitive and stereotyped interests and play

Source : Early diagnosis and course in ASD – IMFAR 2013 – Rebecca Landa

Red flags by 14 months
Low social responsiveness and reciprocity
Infrequent initiation of joint attention and response to social cues
Infrequent playful imitation
Language delay
Repetitive and stereotyped interests and play

Source : Early diagnosis and course in ASD – IMFAR 2013 – Rebecca Landa

Joint Attention

Development psychology – JA is used to encompass behaviours young children use to actively coordinate their attention and interest in shared objects and events with their social partner. [Chris Moore and Philip J Dunham [1]]

(a) Initiate joint attention by alternating gaze - child looking at the toy and person

(b) Respond to joint attention

Source : ESCS

Joint Engagement

Development psychology - More extended periods of shared attention between a child and (typically) their caregiver in which both partners are focused on their social partner and on the shared activity. [Roger Bakeman and Lauren B. Adamson [2]]

Developing computational model of engagement result in determining child’s JA in a social interaction. This will provide metrics for early diagnosis of ASD.

Related Work

1. Human-Robot Interactions
   - Head and hand poses

2. Adult-Child Interactions
   - Object, head trajectories and audio features
   - Body gestures from depth images
   - Electrodermal activity (EDA) of the children from wearable sensors

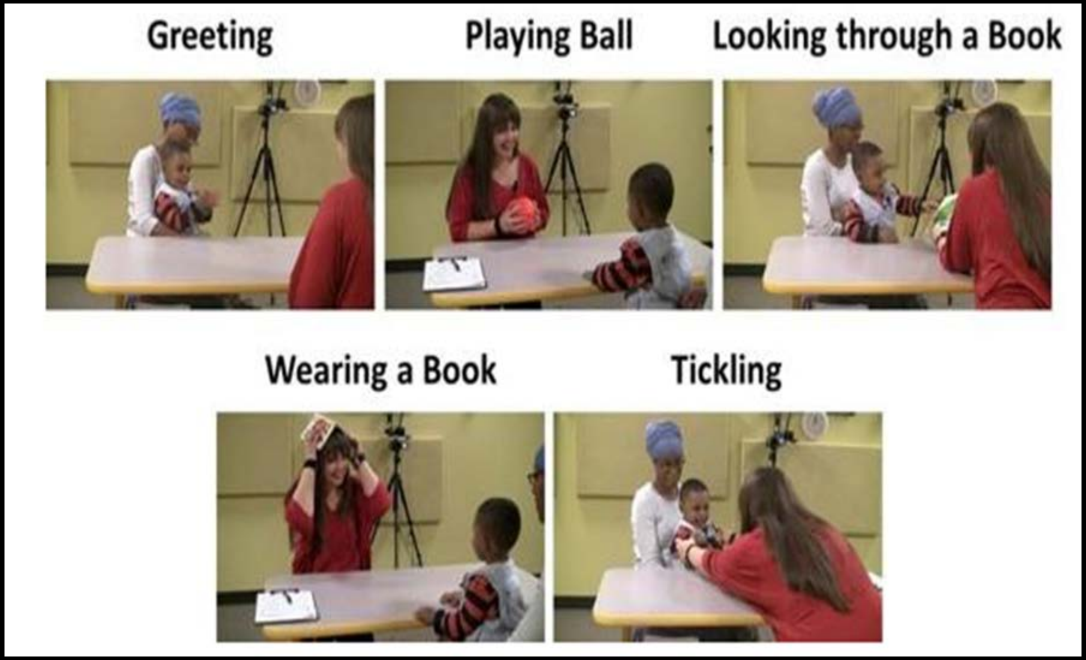
3. Virtual Humans

Computational Approach to Engagement Prediction

1. An investigation of the feasibility of using low-level features in the absence of robust, accurate high-level non-verbal features.

2. A two-stage HCRF + SVM model for engagement prediction using the learned hidden structures of the behaviours.

Multimodal Dyadic Behaviour Dataset (MMDB) [7]



1. 160 sessions
2. 3-5 min semi-structured play between adult and examiner
3. 5 stages

0 = Easy to Engage,
1 = Requires Some Effort to Engage
2 = Requires Extensive Effort to Engage

Example from MMDB dataset - Challenges in automatic extraction of high-level features (head pose)

	Child actions
1	Child looking down
2	Child turning towards other parts of the room
3	Child climbing up on the mother and looking at her
4	Child walking away and coming back after some time
5	Child’s head hair occluding the face
6	Child’s hand occluding the face
7	Child looking down at the book
8	Child’s drinking from a glass occluding the face

These challenges warrant exploring low-level features

Investigation of low-level features

1. When a child dominates in a social interaction, the motion cues reflect the child’s behaviour.

2. The sub-actions leading to a behaviour can be learnt using Hidden Conditional Random Fields (HCRF).

Investigation of low-level features using MMDB

- The released ground truth annotations are used for baseline comparison.
- Compute Behaviour Words using STIP-HOG-HOF features. These are used as low-level features.
- Detect manually guided head pose orientations for each frame. These are used as high-level features.
- Learn a discriminative model using Hidden Conditional Random Fields (HCRF).

Engagement prediction	Ground Truth	Manually guided high-level features	Behaviour words based low-level features
Classification Accuracy	71.2%	69.3%	70.3%

Proposed two-stage algorithm for engagement prediction

Stage 1 – Learning the hidden structures of a behaviour

Input : Behaviour Words
Output : Estimated hidden state probabilities

Behaviour words used as node observations to train a HCRF model

Stage 2 – Training a model using the learnt hidden structures

Input : Estimated hidden state probabilities
Output : Classification Accuracy

Train a SVM using estimated hidden state probabilities as a feature vector

Computing Behaviour Words

1. Detect child’s upper body region

2. Construct a visual vocabulary using Histogram of Optical Flow (HOF) features of all frames using clustering. The cluster centres represent visual words.

3. Map HOFs of each frame to nearest visual word and choose the visual word containing maximum HOF as the frame descriptor.

4. Perform clustering on the frame descriptors and the cluster centres are referred as “Behaviour Words”.

SVM feature computation

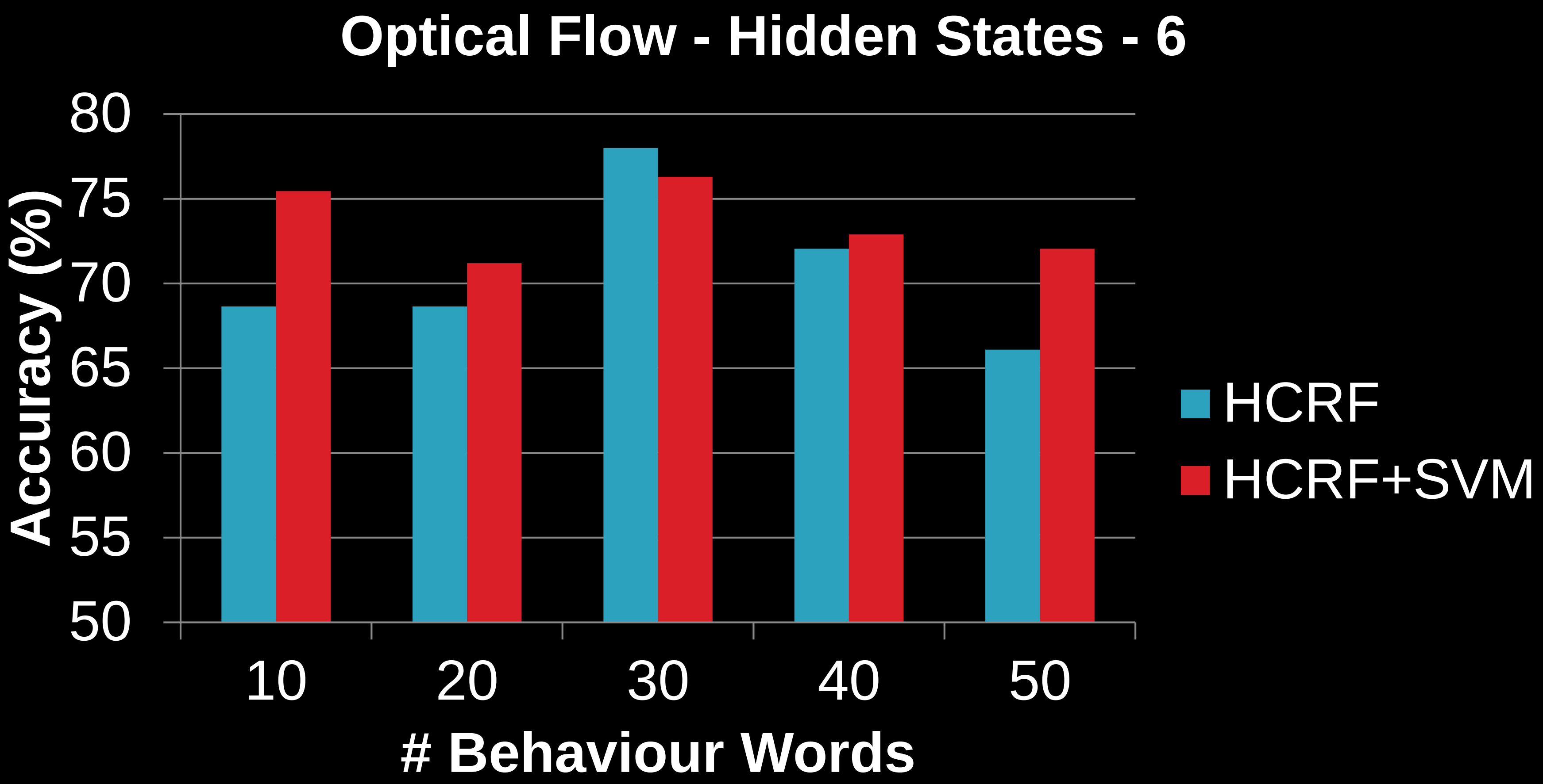
1. The probabilities of each hidden state for each class and each node is available at the end of the training procedure from stage 1.

2. In stage 2, for each hidden state, compute the marginal probability across all classes.

3. The marginal probability of all the hidden states are fused together to form a feature vector.

4. Use this feature vector for training the SVM

Experiments and Results - MMDB



Experiments and Results – J-HMDB

Hidden States	HCRF (A)	A + SVM	A + FV
4	25.0%	28.0%	65.3%
7	30.8%	35.7%	65.5%
10	36.1%	39.0%	66.4%
12	37.5%	42.1%	66.1%
15	35.7%	37.0%	65.0%

Jhuang et al. [10]	57.6%
Fisher Vector (FV) [11]	62.8%

Given that high-level features are difficult to obtain, the low-level features are good substitute for engagement prediction. Two-stage approach results in providing additional complimentary information.

Summary

1. When a child dominates in a social interaction, the motion cues reflect the child’s behaviour

2. The motion based low-level features can be a good substitute for hard to obtain high-level features

3. The learned hidden structures are used in two-stage framework for better recognition performance

4. Future work involve using multi-view learning to exploit the relationship between multimodal signals.

Thank You

References

1. Chris Moore and Philip J Dunham, Joint attention: Its origins and role in development, Lawrence Erlbaum Associates Publishers, Hillsdale, New Jersey and Hove, UK, 1995.

2. Roger Bakeman and Lauren B. Adamson, “Coordinating Attention to People and Objects in Mother-Infant and Peer-Infant Interaction,” Child Development, vol. 55, pp. 1278–1289, Aug. 1984.

3. A.Vinciarelli, M.Pantic, D.Heylen, C.Pelachaud, I.Poggi, F.D’Errico and M.Schroeder, “Bridging the Gap Between Social Animal and Unsocial Machine: A Survey of Social Signal Processing“, IEEE Transactions on Affective Computing, Vol. 3, No. 1, pp. 69-87 (2012).

4. A.Vinciarelli, M.Pantic and H.Bourlard, “Social Signal Processing: Survey of an Emerging Domain“, Image and Vision Computing Journal, Vol. 27, No. 12, pp. 1743-1759 (2009).

5. D. Gatica-Perez, Automatic nonverbal analysis of social interaction in small groups: A review, Image Vis. Comput. (2009), doi:10.1016/j.imavis.2009.01.004

References

5. Michael Tomasello, Chris Moore (Ed), and Philip J. Dunham (Ed). Joint attention as social cognition. Lawrence Erlbaum Associates, Inc,, Hillsdale, NJ, England, 1995.

6. S.S. Rajagopalan, A. Dhall and R. Goecke. Self-Stimulatory Behaviours in the Wild for Autism Diagnosis. *Proceedings of the IEEE International Conference on Computer Vision Workshops, IEEE Workshop on Decoding Subtle Cues from Social Interactions*, Sydney, Australia, 8 Dec 2013.

7. James M Rehg, Gregory D Abowd, Agata Rozga, Mario Romero, Mark A Clements, Stan Sclaroff, and et al. Decoding Children's Social Behavior. In IEEE CVPR, pages 3414–3421, 2013.

8. S.S. Rajagopalan and R. Goecke. Detecting Self-Stimulatory Behaviours for Autism Diagnosis**.** *Proceedings of the International Conference on Image Processing (ICIP 2014)*, Paris, France, 27-30 Oct 2014. (Accepted)

9. Landa RJ1, Gross AL, Stuart EA, Faherty A., Developmental trajectories in children with and without autism spectrum disorders: the first 3 years., Child Dev. 2013 Mar-Apr;84(2):429-42.

References

10. Hueihan Jhuang, Juergen Gall, Silvia Zuffi, Cordelia Schmid, and Michael J. Black, “Towards understanding action recognition,” in International Conference on Computer Vision, 2013

11. Xiaojiang Peng, Changqing Zou, Yu Qiao, and Qiang Peng, “Action recognition with stacked fisher vectors,” in European Conference on Computer Vision, 2014, pp. 581–595..

12. Guido Pusiol, Laura Soriano, Li Fei-Fei, and Michael C. Frank, “Discovering the signatures of joint attention in child-caregiver interaction,” in The annual meeting of the Cognitive Science society CogSci 2014, 2014.

BACKUP

UNIVERSITY OF CANBERRA

Effectiveness of the two-stage approach

Hidden states < optimal hidden states – 2-stage is useful

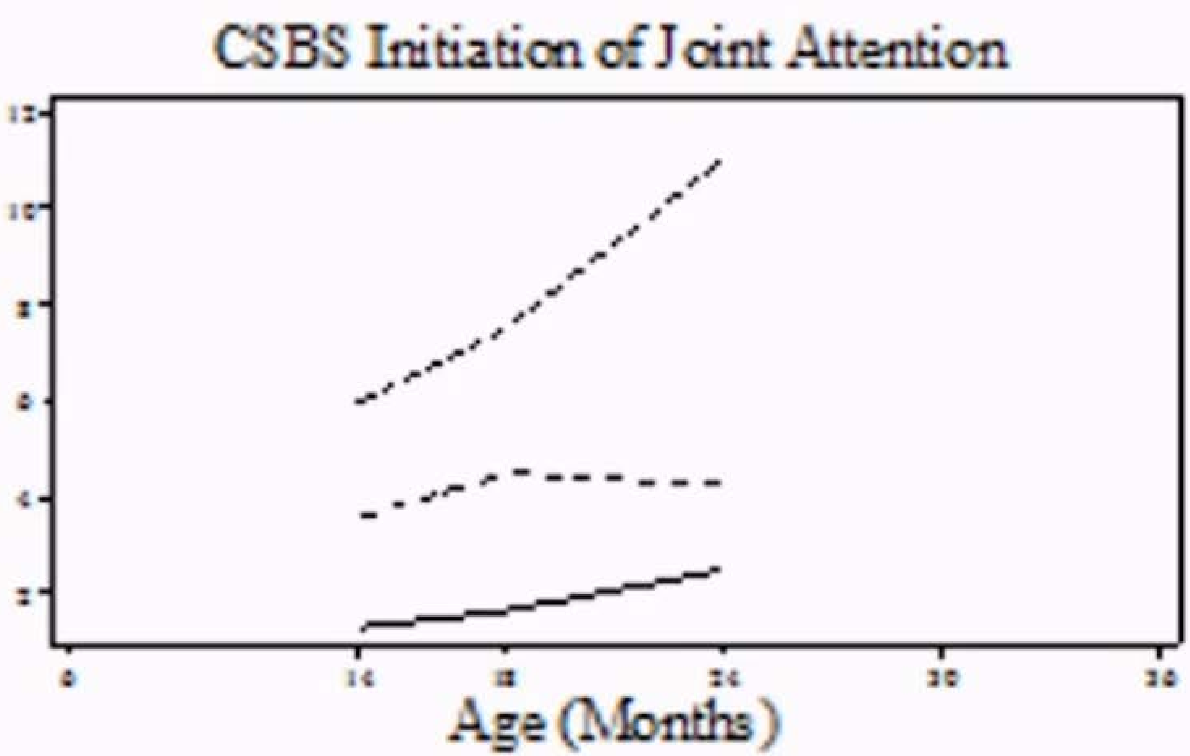
Hidden states > optimal hidden states – 2-stage is useful under certain conditions

Hidden states = optimal hidden states – 2-stage is not needed

Computers will become “Socially Intelligent”, when they can recognize these human social behaviours.

Joint Attention

Child initiates somebody to look at something interesting for the purpose of social sharing



Landa, Gross, Stuart, Faherty. 2013. Child Development

Center for Autism and Related Disorders
at Kennedy Krieger Institute

ASD Atypical Behaviours

Bringing the Early Signs of Autism Spectrum Disorders Into Focus
http://www.youtube.com/watch?v=YtvP5A5OHpU (00:00 – 00:45, 04:24 – 07:56)

Research focus

Developing computational models for machine understanding of human behaviours using non-verbal multimodal cues

Technology that assists clinicians in diagnosing Autism Spectrum Disorders (ASD)