Tactile interaction and social touch: Classifying human touch using a soft tactile sensor

Jiong Sun
University of Skövde
Skövde, Sweden
jiong.sun@his.se

Sergey Redyuk
University of Skövde
Skövde, Sweden
b16serre@student.his.se

Erik Billing
University of Skövde
Skövde, Sweden
erik.billing@his.se

Dan Högberg
University of Skövde
Skövde, Sweden
dan.hogberg@his.se

Paul Hemeren
University of Skövde
Skövde, Sweden
paul.hemeren@his.se

ABSTRACT
This poster presents an ongoing study on affective human-robot interaction. In our previous research, touch type is shown to be informative for communicated emotion. Here, a soft matrix array sensor is used to capture the tactile interaction between human and robot and 6 machine learning methods including CNN, RNN and C3D are implemented to classify different touch types, constituting a pre-stage to recognizing emotional tactile interaction. Results show an average recognition rate of 95% by C3D for classified touch types, which provide stable classification results for developing social touch technology.

Author Keywords
Tactile interaction, social touch, affective HRI, machine learning

INTRODUCTION
Social touch is important to the physical interactions between humans [1]. Communicating affect through social touch can increase social acceptance and bonding. While many types of emotions are communicated using other modalities than touch, e.g., face expressions, touch appears to be the preferred channel for communication of intimate emotions and critical for social bonding [2]. Affective tactile interaction is therefore likely to be critical for successful applications of social robotics.

Currently, social touch technology is being used to develop a more human-like interaction with robots. The goal is to achieve a greater social acceptance of the robots by humans [1] and to offer enhanced human-robot interaction functionality. This technology should therefore support different kinds of tactile/touch interactions between people and physically embodied artificial social agents. The further purpose of this technology is to decode emotions that are expressed in human-robot interaction scenarios. The signals from the tactile interaction can also be used to complement auditory and visual sensory information to create a more robust classification of social touch [3].

Based on previous results from a human-robot interaction study [4], we selected seven types of touches (stroke, scratch, poke, press, push, squeeze and grab) and used a soft array sensor to classify them. The results from this previous study showed that squeezing, stroking and pressing constituted more than half (59%) of all tactile interaction in the study.

Six different machine learning techniques have been applied in our experiments for touch classifications. The purpose of comparing these machine learning techniques is to determine the empirical extent of classification difference between the techniques in terms of accuracy and the possible confusions between different touches. Ninety-five percent of accuracy has been achieved by using C3D (3D convolutional neural networks).

EXPERIMENTAL SETUP
An FSR (Force Sensitive Resistor) matrix array sensor MS9724 together with an Arduino Leonardo compatible board Snowboard from Kitronyx [5] are used, as shown in Fig 1. The matrix sensor has 160 force sensing nodes (16 rows x 10 columns) with 128 mm x 80 mm active sensing area and is designed for true force and multi-touch detection. The working sampling rate of the sensor is 30Hz. The matrix sensor is placed on a Pepper [6] robot's arm, as shown in Fig.1.

Seven different types of touch movements are adopted in the experiment: stroke, scratch, poke, press, push, squeeze and grab, as shown in Fig 2. The length of the movement is 60 samples, which equals to 2 seconds. Table 1 shows descriptive statistics of the training set. Grab and Push have the largest number of pressed nodes and magnitude. Poke and Scratch have the least number of nodes activated, while Scratch has a smaller magnitude.

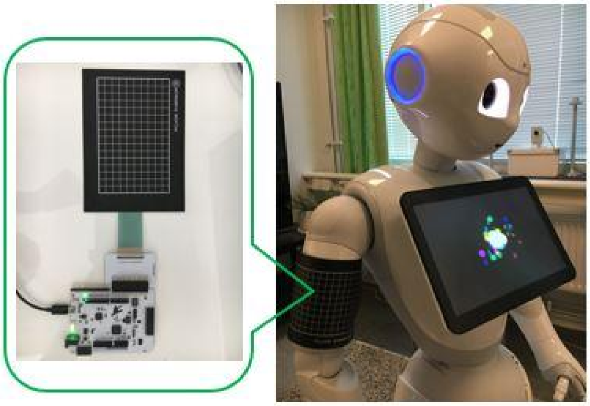

Figure 1. The Kitronyx matrix array sensor and the Snowboard microcontroller on Pepper's arm.

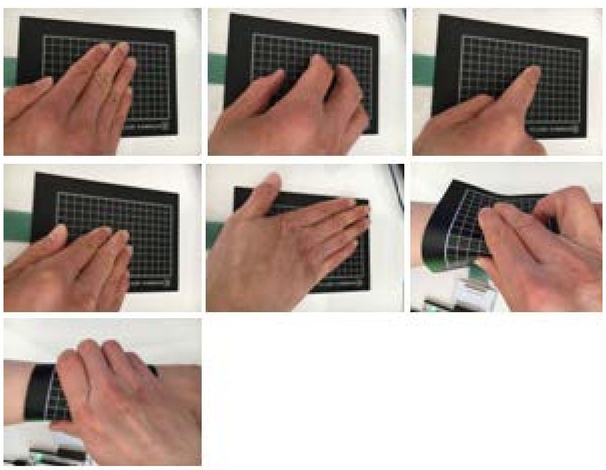

Figure 2. 7 types of touch movements adopted in the experiment. Row 1, from left to right: stroke, scratch, poke; Row 2: press, push, squeeze, Row 3: grab.

Touch type	Active nodes	Magnitude	Avg. magnitude	Duration (samples@30 Hz)
Stroke	7-28	42-367	6-13	44-60
Scratch	2-4	16-55	10-18	21-46
Poke	1-4	27-114	15-55	20-43
Press	10-20	131-364	12-22	30-60
Push	17-41	183-438	7-13	32-60
Squeeze	10-18	102-210	9-14	35-51
Grab	17-39	164-477	6-12	32-60

Table 1. Characteristics of different touches

CLASSIFICATION MODELS

Six machine learning techniques have been evaluated: 1) Support Vector Machines (SVM), 2) Decision Trees (DT), 3) Convolutional Neural Network (CNN), 4) Recurrent Neural Network (RNN), 5) Long-term Recurrent Convolutional Neural Network (RCNN), and 6) 3-dimensional Convolutional Network (C3D).

Support Vector Machines

An SVM with a polynomial kernel function of degree 3 was used.

Convolutional Neural Networks

The CNN model used has two parts, consisting of convolutional and dense layers. The first part comprise three convolutional layers contain 32, 64 and 128 kernels respectively. Then, there is a dropout layer after each convolutional layer to prevent overfitting. The second part of the network has two hidden dense layers of size 128 and 64 and ReLU activation function followed by the output layer with a softmax activation function. This model consists of 756,871 parameters to optimize.

Recurrent Neural Networks

The RNN model contains three LSTM units and is trained to classify a sequence of 5 neighboring frames. A dropout layer follows each LSTM layer to prevent overfitting. Then comes one hidden layer of size 64 with ReLU activation following by the output layer with a softmax activation function. This model consists of 627,079 parameters to optimize.

Long-term Recurrent Convolutional Neural Networks

RCNN model is a combination of recurrent and convolutional neural networks to capture both spatial and temporal features. A model proposed in [7] takes an input data and processes it by using CNN units to learn spatial features and then passes these extracted features to LSTM units for sequence learning.

This model consists of three time distributed convolutional layers of sizes 32, 64 and 128. The second part of the model contains three LSTM units to handle temporal features. Then comes one hidden layer of size 64 following by the output layer with a softmax activation function. ReLU activation is used for each layer. A dropout layer follows each convolutional and LSTM layers to prevent overfitting. This model has 3,894,151 parameters to optimize.

3-Dimensional Convolutional Neural Networks

In regular CNNs, convolutions are applied on the two dimensional feature maps to compute features which can be shared in space. When applied to sequence analysis, it is desirable to capture the motion information encoded in multiple contiguous frames. 3D convolutions for CNNs proposed in [8] computes features from both spatial and temporal dimensions. The 3D convolution is achieved by convolving a 3D kernel to the cube formed by stacking multiple contiguous frames together.

The C3D model used for this research contains three 3D convolutional layers with the same amount of kernels per layer - 32, 64 and 128 kernels respectively -

as the CNN model has. Then come two hidden dense layers of size 128 and 64 and ReLU activation function followed by the output layer with a softmax activation function. This model contains 13,393,607 parameters to optimize.

TRAINING PROCESS

All the models are divided into two groups. SVM, Decision Trees, Random Forests, Extremely Randomized Trees and CNN treat each row (frame) as a separate example. Other models - RNN, RCNN and C3D - analyze sequences of frames of size 5 to predict the type of touch.

Frame sequences are extracted without overlapping to make sure that none of the training set examples appear in the test set. Another strategy to allow overlapping but separate files for training and test sets was denied because cross-validation was used for the solution evaluation stage which required one general dataset instead of training and test sets. As a drawback of this solution, the number of examples decreased drastically.

For the models that classify separate rows, 10268 examples are used - 9241 examples as a training set and 1027 examples as a test set. Other models that deal with sequences, 1964 examples are used - 1768 examples as a training set and 198 examples as a test set.

The number of epochs used for training is rather limited as the training process for deep learning models is time-consuming. This number should be extended to train the production system.

All computations for neural networks are performed on an HP Z420 Workstation with a single GPU unit Nvidia M4000.

EXPERIMENTAL RESULTS

K-fold cross-validation was used to evaluate models performance. Initial training and test sets were concatenated together and then split into 10 chunks. Table 2 shows the performance metrics for each implemented model and Figure 3 – 8 show the confusion matrices for the 6 classification methods.

Model	Epochs	Accuracy
SVM	-	84.35%
DT	-	85.71%
CNN	100	82.29%
CNN	1000	85.42%
RNN	10	91.09%
RNN	20	91.70%
RCNN	10	91.91%
RCNN	20	93.02%
C3D	10	93.48%
C3D	20	**94.76%**

Table 2. Performance metrics for different models

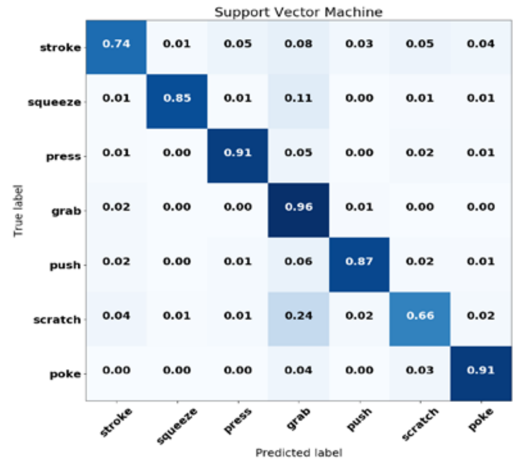

Figure 3. Confusion matrix for SVN

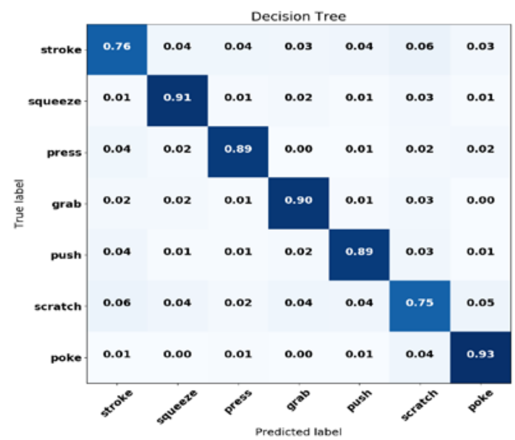

Figure 4. Confusion matrix for DT

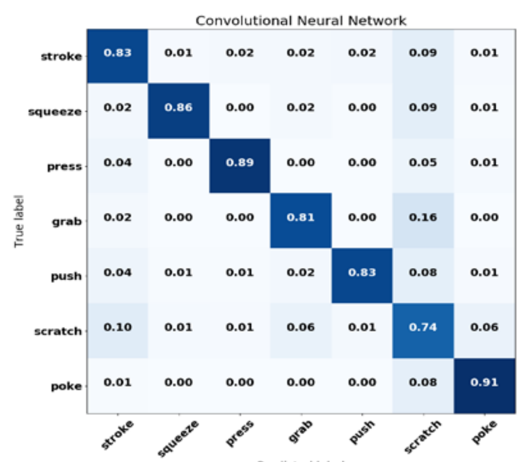

Figure 5. Confusion matrix for CNN with epoch size 1000

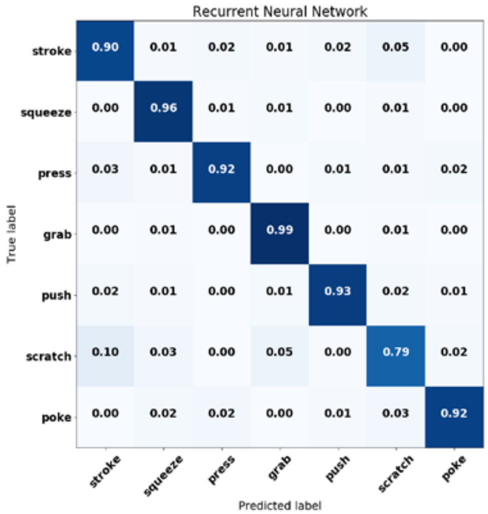

Figure 6. Confusion matrix for RNN with epoch size 20

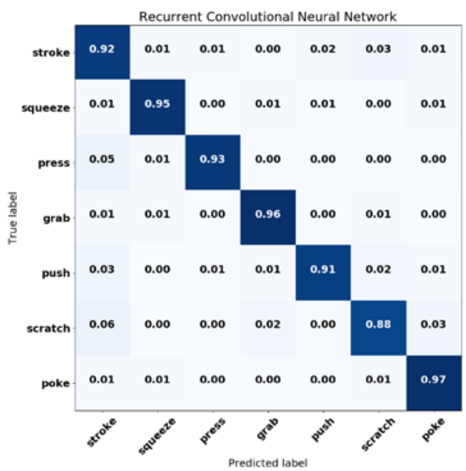

Figure 7. Confusion matrix for RCNN with epoch size 20

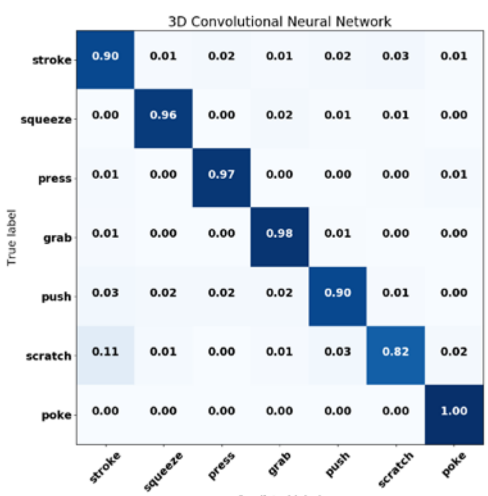

Figure 8. Confusion matrix for C3D with epoch size 20

Seven most commonly used touches that occur in the expression of different emotions in human-robot interaction were selected to illustrate how touch categorization could be achieved by using input from a soft array sensor to categorize different types of touch. Six different classification methods (SVM, DT, CNN, RNN, RCNN, C3D) were implemented. Experimental results show that C3D has the highest average recognition rate of 94.8% in all 6 classifiers. The recognition rates of the 7 touches in C3D range from 82% to 100%, among which scratch has the lowest rate of 82% and the rest are above 90%, while in other methods scratch also has the lowest recognition rate, ranging from 66% (SVM) to 88% (RCNN).

CONCLUSIONS

The variability in the recognition rate between different touch types could be understood as some touch types, e.g. squeeze and push, land close to each other in a feature space, reducing the recognition rate, while other touch types, e.g., stroke and poke, appears well-separated from other types of touch. The contribution that this research makes to the field of human-robot interaction is the evaluation of classification performance by machine learning techniques of seven commonly used touches to communicate emotions between people.

One interesting connection regarding the results could be made to human-human tactile interaction, where different types of touch have been shown to carry varying degrees of emotional information [8]. For example, push is indicative for anti-social emotions such as anger and disgust, while other touch types, e.g., squeeze, appears when communicating a variety of emotions and as a consequence carries less information. We therefore hypothesize that the more similar two emotions are to each other, the less distinct communicated touches are from each other, and therefore more difficult to classify. This is in line with conceptual space theory [9], predicting that the distance between categories, such as different emotions, should be reflected in their kinematic expressions. The next step in this research is to conduct emotion-based human-robot interactions using the matrix array sensor (or a similar device). This would allow us to analyse directly the connection between the physical tactile interaction and the communicated emotions.

ACKNOWLEDGMENTS

This work has been carried out as part of the project "Design, Textiles and Sustainable Development" funded by Region Västra Götland (VGR) funding agency.

REFERENCES

1. Gijs Huisman. 2017. Social touch technology: A survey of haptic technology for social touch. *IEEE Transactions on Haptics* (09 January).
2. Betsy App, Daniel N. McIntosh, Catherine L. Reed, and Matthew J. Hertenstein. 2011. Nonverbal channel use in communication of emotion: how may depend on why. *Emotion*, 11(3), 603–617.
3. Takato Horii, Yukie Nagai, and Minoru Asada. 2013. Touch and emotion: Modeling of developmental differentiation of emotion lead by tactile dominance. In *Proceedings of the 3rd IEEE International Conference on Development and Learning and on Epigenetic Robotics*. August.
4. Rebecca Andreasson, Beatrice Alenljung, Erik Billing, and Robert Lowe. 2017. Affective Touch in Human-Robot Interaction: Conveying Emotion to the Nao Robot. Submitted to *International Journal of Social Robotics*.
5. Kitronyx Inc. http://www.kitronyx.com
6. Pepper robot, https://www.ald.softbankrobotics.com/en/ cool-robots
7. Jeff Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. 2015 Long-term Recurrent Convolutional Networks for Visual Recognition and Description. *IEEE Conference on Computer Vision and Pattern Recognition*. 2625-2634.
8. Shuiwang Ji, Wei Xu, Ming Yang, and Kai Yu. 3D convolutional neural networks for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 221-231.
9. Peter Gärdenfors, and Mary -Anne Williams 2001. Reasoning about categories in conceptual spaces. In *IJCAI*, (August) 385-392.