Detecting Conflict in Couples

Aditya Gujral
Texas A&M University
Email: aditya.gujral@tamu.edu

***Abstract*—With the increase in wearable and mobile technology, more data for individuals is available. This data can potentially be used to improve individual lifestyle and self-care. The results of this project demonstrate that it is possible to detect conflict in couples using wearable technology and mobile data in an uncontrolled environment. This can have multiple practical applications, including digital intervention by recommendation of alternate behavior.**

***Keywords*—Machine Learning, Wearable Technology, Conflict Detection, SVM, LSTM.**

I. Introduction

Over the past decade there has been enormous growth in mobile and wearable technology [1][2]. Wearable systems allow unobtrusive monitoring over extensive periods of time. Digital health interventions (DHIs) are being created to support patients in their everyday life for improved self health-care [3]. Such interventions can significantly improve self-management of various health conditions by enabling continuous insight into data collected.

Various mobile and wearable devices such as smart phones, sleep monitors, activity trackers and medical devices for measuring blood pressure are now more accessible. The prices of such devices are also becoming affordable with increased research, demand and competition. The ability of such devices to capture data allow for many applications. To keep up with the enormous growth of such data, novel methods to process and interpret the data are extremely important.

Primary concerns of individuals include security and privacy of such data [5][6]. All approaches must consider these issues and also respect existing any regulations regarding the same. Ethical considerations should hence be be a primary concern of such applications.

There are many possible applications of such data, and this project explores one such application - detecting conflict in couples. Conflict between couples can have profound impact on individual well-being [7]. Conflict in romantic relationships can lead to depression and other psychological problems [8]. Escalation of such conflict, particularly divorce, can lead to a downward trajectory from which individuals may never recover [9]. The quality of the relationship affect both mental and physical health. Detecting conflict in uncontrolled environment can have multiple practical applications, including the development of online automated intervention systems for suggesting an alternative behavior to avoid conflict.

II. Related Work

Previous studies that attempt to detect conflict have usually been attempted in controlled environments [10][11]. It is difficult to apply such methods in uncontrolled environments because of the assumptions about the controlled behaviour. Moreover, these works focus on unimodal or bimodal features, whereas this project considers Multimodal features. Interaction between couples can be better captured using Multimodal feature detection in couples. It is also important to consider that the data collected in real-world scenarios can be noisy. Other challenges such as time synchronization between devices should also be taken into account. Furthermore, fluctuations in features, such as Electrocardiography, can be due to exercise or anxiety apart from the event of interest [12].

Previous research by Timmons et al. [4] worked on the same data-set. The research includes detailed analysis of the data-set and uses autoeoncder stacked with decision trees for classification. This project aims to extend the work by trying other models that could better classify the data.

III. Data description

The data was provided through the courtesy of the USC Couple Mobile Sensing Project (homedata.github.io). Wearable technology, such as ambulatory physiological monitor, was used to measure electrocardiography, electrodermal activity, body temperature, heart rate variability, skin conductance level and and the frequency of skin conductance response. Smart mobiles were used to collect GPS data throughout the day, which can be used to calculate physical distance between the couple. Audio was recorded using smart phone. Six minute of total audio was recorded every twelve minutes, from 10 am till bedtime [4]. Linguistic Inquiry and Word Count (LIWC) software [13] was used for Language and acoustic feature extraction. Each sample in the final data-set was at the interval of 3 minutes.

The data was pre-processed for missing values. Features with all missing values were dropped. Samples for which output label was missing were dropped. Remaining missing values were replaced by feature mean using the mean imputation (MI) method [14].

IV. Proposed Solutions

All the features were scaled to [0,1] range. The data-set is imbalanced as conflict instances contributed for only about 5% of the data. Any classifier that always predicts no-conflict will have approximately 95% accuracy. Clearly, accuracy is hence not a good measure. Instead, using precision, recall and F1-score for each class gives a good measure of the model.

The data is split into testing (20%) and training data (80%). The training data is further used for 15-fold stratified cross-validation to estimate the hyper-parameters of each model.

Original Data

Couple ID	Time	Conlict label	Feature 1	Feature 2	...
1	10:30	0	0.2	0.5	...
1	10:33	0	0.3	0.6	...
1	10:36	1	0.4	0.7	...
1	10:39	0	0.1	0.4	...

Sequenced data for LSTM

Sequence Id	Couple ID	Time	Conlict label	Feature 1	Feature 2	...
1	1	10:30	0	0.2	0.5	...
	1	10:33		0.3	0.6	...
2	1	10:33	1	0.3	0.6	...
	1	10:36		0.4	0.7	...
3	1	10:36	0	0.4	0.7	...
	1	10:39		0.1	0.4	...

Fig. 1. Converting data to sequential form for LSTM

Stratified cross-validation is necessary to ensure that the proportion of each class over folds is comparable. After the hyper-parameters for each model are determined, the results are evaluated on the untouched testing data.

Support Vector Machines, Feedforward Neural Network and Long Term Short Memory (LSTM) networks are used for analysis. The motivation for using LSTM is to capture sequential information in the data. While Hidden Markov Models can also be used to capture sequential information, Deep Neural Networks, using LSTM, capture contextual information more flexibly than decision trees used in the HMM-based system [15]. LSTM also overcomes the limitation of vanishing gradients in traditional Recurrent Neural Networks [16].

For LSTM, samples at time t-1 and time t formed one sequence. Hence each sequence is of length 2, as shown in Fig. 1. Note that the sequence is formed after splitting the data, to ensure there is no contamination of testing/validation data.

To ensure a higher recall rate for conflict, misclassification of conflict is penalized more by using a weighted class approach in all the models. To choose the hyper-parameters, value of 0.5*(recall of no-conflict + recall of conflict) is used. The weight for each class was again a hyper-parameter. Auto-encoder was tried for feature reduction but no significant improvement was seen for any model, hence it is not included in the results. All the hyper-parameters, such as number of nodes in each layer, activation function, dropout and number of layers for LSTM and Feedforward Neural Network and parameter C and kernel for SVM, are calculated using 15-fold stratified cross-validation as explained earlier.

The Feedforward Neural Network contains the input layer, one hidden layer and one output sigmoid node. In the LSTM network, the first layer has LSTM nodes, the second layer has feedforward nodes and the output layer is a single sigmoid node. Activation functions for input and hidden layers is determined to be ReLu based on cross-validation for both the networks.

As we are trying to improve the recall, a custom error (0.5*(recall of no-conflict + recall of conflict)) for back-propagation, based on recall instead of accuracy, was tried of both LSTM and Feedforward Neural Network. But there was no improvement in results and hence the approach was discarded.

A. *Results and Limitations*

TABLE I. RESULTS FOR SVM ON TESTING DATA

Class	Precision	Recall	F1 - score	Support
No - Conflict	0.98	0.84	0.90	209
Conflict	0.19	0.73	0.30	11
Average	0.94	0.83	0.87	220
Unweighted Average	0.59	0.79	0.6	

TABLE II. RESULTS FOR FEEDFORWARD NEURAL NETWORK ON TESTING DATA

Class	Precision	Recall	F1 - score	Support
No - Conflict	0.96	0.76	0.85	209
Conflict	0.12	0.42	0.19	11
Average	0.92	0.81	0.85	220
Unweighted Average	0.54	0.59	0.52	

TABLE III. RESULTS FOR LSTM ON TESTING DATA

Class	Precision	Recall	F1 - score	Support
No - Conflict	0.99	0.73	0.84	209
Conflict	0.15	0.91	0.26	11
Average	0.95	0.74	0.81	220
Unweighted Average	0.57	0.83	0.55	

SVM has the highest precision for conflict on the testing data. LSTM performs comparable and has the highest recall for conflict. Feedforward Neural Network however do not perform well and is not suitable for the application. Having a high recall rate for LSTM and SVM models enables the opportunity to suggest alternate behavior during conflict. Low precision for conflict is however a concern. As the representation of conflict class is very low, overall precision and fraction of time we report false positive for conflict is still acceptable. The unweighted average (UA) is the average value of measure over each class. For example, the unweighted average for recall is 0.5*(recall of no-conflict + recall of conflict).

V. CONCLUSION AND FUTURE WORK

The analysis supports the claim that complex psychological states can be detected in an uncontrolled environment using multimodal data. Recommendations for alternate behaviour prompts to decrease likelihood conflicts could be explored as future work. These interventions can possibly include personal informatics that enable individuals to engage in self analysis.

The project uses class weight to account for imbalanced data-set. Under-sampling techniques such as random majority under-sampling with replacement and Neighborhood Cleaning Rule [17] can be used. Over-sampling techniques such as random minority over-sampling with replacement and Synthetic Minority Over-sampling Technique [18] can also possibly give better results.

Personalization of model for each couple can be considered to improve the results. The LSTM sequence is created by stacking samples at time t-1 and t. Various depth of such stacking can be experimented with.

A. Takeaway messages (based on personal opinion)

- SVM usually performs very well with minimal tuning as they have fewer hyper-parameters (relative to Deep Neural Networks). Hence, they provide a good baseline for comparison.
- Start simple!
- If Data is scarce, complex models usually will not work so good.
- Do not ever touch the testing data, ever. Just use the testing data for final results only. This is among the golden rules for machine learning.

ACKNOWLEDGMENT

I would like to thank the continuous guidance and support of my professor, Theodora Chaspari. I would also like to thank the University of Southern California (USC) Couple Mobile Sensing Project (homedata.github.io) for providing the dataset.

REFERENCES

[1] Katrin Hansel, Natalie Wilde, Hamed Haddadi, Akram Alomainy, *Challenges with Current Wearable Technology in Monitoring Health Data and Providing Positive Behavioural Support*, Proceedings of the 5th EAI International Conference on Wireless Mobile Communication and Healthcare, p.158-161, October 14-16, 2015, London, Great Britain

[2] X. F. Teng, Y. T. Zhang, C. C. Y. Poon, and P. Bonato, *Wearable medical systems for p-health*, IEEE Rev. Biomed. Eng., vol. 1, pp. 6274, 2008.

[3] Liang, X., Q. Wang, X. Yang, J. Cao, J. Chen, X. Mo, J. Huang, L. Wang and D. Gu, *Effect of mobile phone intervention for diabetes on glycaemic control: a meta-analysis*, Diabetic medicine 28 (4), 455 - 463 2011.

[4] Adela C. Timmons, Theodora Chaspari, Sohyun C. Han, Laura Perrone, Shrikanth S. Narayanan, and Gayla Margolin, *Using Multimodal Wearable Technology to Detect Conflict among Couples*, Computer, vol. 50, pp. 50-59, 2017, ISSN 0018-9162.

[5] B. Yuksel, A. Kupcu, and O. Ozkasap, *Research issues for privacy and security of electronic health services*, Future Generation Computer Systems, Vol. 68, pp. 113, 2017.

[6] Nabil Sultan, *Reflective thoughts on the potential and challenges of wearable technology for healthcare provision and medical education*, International Journal of Information Management Volume 35, Issue 5, Pages 521-526, October 2015.

[7] Coyne JC, Downey G, *Social factors and psychopathology: stress, social support, and coping processes* Annu. Rev. Psychol. 42:401-25, 1991.

[8] Tesser A, Beach, *Life events, relationship quality, and depression: an investigation of judgement discontinuity*, J. Pers. Soc. Psychol. 74:36-52, 1998.

[9] P.R. Amato, *The Consequences of Divorce for Adults and Children*, J. Marriage and Family, vol. 62, no. 4, pp. 12691287, 2000.

[10] S. Kim, F. Valente, and A. Vinciarelli, *Automatic Detection of Conflict in Spoken Conversations: Ratings and Analysis of Broadcast Political Debates*, Proc. IEEE Intl Conf. Acoustic, Speech, and Signal Processing (ICASSP 12), 2012.

[11] H. Hung and G. Englebienne, *Systematic Evaluation of Social Behavior Modeling with a Single Accelerometer* Proc. ACM Intl Joint Conf. Pervasive and Ubiquitous Computing (UbiComp 13), pp. 127130, 2013.

[12] Detry J-MR, Kapita BM, Cosyns J, et al, *Diagnostic Value of History and Maximal Exercise Electrocardiography Men and Women Suspected of Coronary Heart Disease* Circulation 56:756-769, 1977

[13] J.W. Pennebaker et al., *The Development and Psychometric Properties of LIWC2007*, 2007; www.liwc.net/LIWC2007LanguageManual.pdf.

[14] J.L. Schafer, *Analysis of incomplete multivariate data* , Chapman &Hall, Florida, 1997.

[15] T. Koriyama, T. Kobayashi, *A Comparison of Speech Synthesis Systems Based on GPR, HMM, and DNN with a Small Amount of Training Data*, Proc. INTERSPEECH, pp. 3496-3500, 2015.

[16] Hochreiter, S. and Schmidhuber, J., *Long short-term memory*, Neural Computation 9, 173580 (Nov 1997).

[17] Laurikkala J. *Improving Identification of Difficult Small Classes by Balancing Class Distribution*, Artificial Intelligence in Medicine. AIME 2001.

[18] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, *SMOTE: Synthetic Minority Over-sampling Technique*, Journal of Artificial Intelligence Research (JAIR) 16, 321-357, 2002.