CHAPTER 1

INTRODUCTION

Collaborative learning methods have been implemented broadly by organizations at all stages, as research recommends that active human involvement in cohesive and micro group communications is critical for effective learning. The objective of this work is to propose a machine learning-based methodology system architecture and algorithms to find patterns of learning, interaction, and relationship and effective assessment for a complex system involving massive data that could be obtained from a proposed collaborative learning environment (CLE). Collaborative learning may take place between dyads or larger team members to find solutions for real-time events or problems, and to discuss concepts or interactions during situational judgment tasks (SJT). Modelling a collaborative, networked system that involves multimodal data presents many challenges. This paper focuses on proposing a Machine Learning - (ML)-based system architecture to promote understanding of the behaviours, group dynamics, and interactions in the CLE. Our framework integrates techniques from computational psychometrics (CP) and deep learning models that include the utilization of convolutional neural networks (CNNs) for feature extraction, skill identification, and pattern recognition. Our framework also identifies the behavioural components at a micro level, and can help us model behaviours of a group involved in learning. In current research, an important line of inquiry focuses on finding accurate evidence and valid assessment of these micro-level interactions which supports collaborative learning. Even though there is a long practice of using mathematical models for modelling human behaviour, the author introduced a computational psychometrics-based method for modelling characteristics of real behaviour. Here, the article provides with a way to extract dynamic interaction features from multimodal data for modelling and analysing actual situations. In here, a three-stage method is been proposed to explore and study collaborative group behaviours. The first stage integrates and processes multimodal data obtained in a collaborative learning environment (CLE) that includes sensor input, audio, video, eye tracking, facial expressions, movement, posture, gestures, and behavioural interaction log data. The second stage performs feature extraction and cloud computation using computational psychometrics (CP) and convolutional neural network (CNN)-based deep learning for skill, pattern, and trend identification. Finally, the third stage uses the parameters measured in the previous two stages to understand and model group interactions, competencies, and collaborative behaviour at a micro-level. The third stage uses machine learning for effective assessment and visualization of group dynamics such as correctly assessing the increase in the groups' level of shared understanding of different perspectives, and ability to clarify misconceptions. Modelling a collaborative, networked system that involves multimodal data presents many challenges. This paper focuses on proposing a Machine Learning-(ML)-based system architecture to promote understanding of the behaviours, group dynamics, and interactions in the CLE.

CHAPTER 2

LITERATURE SURVEY

2.1. "ImageNet classification with deep convolutional neural networks", A. Krizhevsky, I. Sutskever, and G. Hinton, 2012.

ImageNet is a dataset of over 15 million labelled high-resolution images belonging to roughly 22,000 categories. The images were collected from the web and labelled by human labellers using Amazon's Mechanical Turk crowd-sourcing tool. Starting in 2010, as part of the Pascal Visual Object Challenge, an annual competition called the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) has been held. ILSVRC uses a subset of ImageNet with roughly 1000 images in each of 1000 categories. In all, there are roughly 1.2 million training images, 50,000 validation images, and 150,000 testing images. Image classification is the process of taking an input (like a picture) and outputting a class (like "cat") or a probability that the input is a particular class ("there's a 90% probability that this input is a cat"). A CNN works by extracting features from images. This eliminates the need for manual feature extraction. The features are not trained. They're learned while the network trains on a set of images. This makes deep learning models extremely accurate for computer vision tasks. CNNs learn feature detection through tens or hundreds of hidden layers. Each layer increases the complexity of the learned features. Instead of feeding the entire image as an array of numbers, the image is broken up into a number of tiles, the machine then tries to predict what each tile is. Finally, the computer tries to predict what's in the picture based on the prediction of all the tiles. This allows the computer to parallelize the operations and detect the object regardless of where it is located in the image.

Advantages: This method identifies every level of visual information is processed and also has the ability to distinguish images based on few test-case scenarios

Disadvantages: This method has a high amount of operations and is unable to maximize inter-class variability.

2.2. "Twitter Sentiment Analysis with Deep Convolutional Neural Networks", Aliaksei Severyn, Alessandro Moschitti, 2015.

Twitter sentiment analysis technology provides the methods to survey public emotion about the events or products related to them. Most of the current researches are focusing on obtaining sentiment features by analysing lexical and syntactic features. These features are expressed explicitly through sentiment words, emoticons, exclamation marks, and so on. In this paper, we introduce a word embedding method obtained by unsupervised learning based on large twitter corpora, this method using latent contextual semantic relationships and co-occurrence statistical characteristics between words in tweets. These word embedding are combined with n-grams features and word sentiment polarity score features to form a sentiment feature set of tweets. The feature set is integrated into a deep convolution neural network for training and predicting sentiment classification labels. We experimentally compare the performance of our model with the baseline model that is a word n-grams model on five Twitter data sets, the results indicate that our model performs better on the accuracy and F1-measure for twitter sentiment classification. Sentiment Analysis, also called opinion mining or emotion AI, is the process of determining whether a piece of writing is positive, negative, or neutral. A common use case for this technology is to discover how people feel about a particular topic. Sentiment analysis is widely applied to reviews and social media for a variety of applications. Sentiment analysis can be performed in many different ways. Many brands and marketers use keyword-based tools that classify data (i.e. social, news, review, blog, etc.) as positive/negative/neutral. Existing approaches to sentiment analysis can be grouped into three main categories: knowledge-based techniques, statistical methods, and hybrid approaches. Knowledge-based techniques classify text by affect categories based on the presence of unambiguous affect words such as happy, sad, afraid, and bored. Some knowledge bases not only list obvious affect words, but also assign arbitrary words a probable "affinity" to particular emotions. Statistical methods leverage elements from machine learning such as latent semantic analysis, support vector machines, "bag of words", "Point wise" for Semantic Orientation, and deep learning. More sophisticated methods try to detect the holder of a sentiment (i.e., the person who maintains that affective state) and the target (i.e., the entity about which the affect is felt). To mine the opinion in context and get the feature about which the speaker has opined, the grammatical relationships of words are used. Grammatical dependency relations are obtained by deep parsing of the text. Hybrid approaches leverage both machine learning and elements from knowledge representation such as ontologies and semantic networks in order to detect semantics that are expressed in a subtle manner, e.g., through the analysis of concepts that do not explicitly convey relevant information, but which are implicitly linked to other concepts that do so. Open source software tools deploy machine learning, statistics, and natural language processing techniques to automate sentiment analysis on large collections of texts, including web pages, online news, internet discussion groups, online reviews, web blogs, and social media. Knowledge-based systems, on the other hand, make use of publicly available resources, to extract the semantic and affective information associated with natural language concepts. Sentiment analysis can also be performed on visual content, i.e., images and videos (see Multimodal sentiment analysis). One of the first approaches in this direction is SentiBank utilizing an adjective noun pair representation of visual content. In addition, the vast majority of sentiment classification approaches rely on the bag-of-words model, which disregards context, grammar and even word order. Approaches that analyses the sentiment based on how words compose the meaning of longer phrases have shown better result,[34] but they incur an additional annotation overhead. A human analysis component is required in sentiment analysis, as automated systems are not able to analyse historical tendencies of the individual commenter, or the platform and are often classified incorrectly in their expressed sentiment. Automation impacts approximately 23% of comments that are correctly classified by humans. However, humans often disagree, and it is argued that the inter-human agreement provides an upper bound that automated sentiment classifiers can eventually reach.

Advantages: It has a strong ability to accept huge number of data, high computational power and faster and efficient performance.

Disadvantages: This method lacks the ability to be spatially invariant to input data.It also has a high computational cost.

2.3. "A. von Davier, "Computational Psychometrics in support of collaborative educational assessments," *Journal of Educational Measurement*, vol. 54, pp. 3-11, 2017."

The behaviour of each individual affects the behaviour of others and, conversely, each one behaves considering the crowd as a whole and the individual others. Modelling behaviour is possible at the micro level through computational neuroscience and at the macro level (society) through computational psychology (e.g., social network analysis and mathematical modelling). Computational Psychometrics is an interdisciplinary field fusing theory-based psychometrics, learning and cognitive sciences, and data-driven AI-based computational models as applied to large-scale/high-dimensional learning, assessment, biometric, or psychological data. Computational psychometrics is frequently concerned with providing actionable and meaningful feedback to individuals based on measurement and analysis of individual differences as they pertain to specific areas of enquiry. Computational psychometrics incorporates both theoretical and applied components ranging from item response theory, classical test theory, and Bayesian approaches to modelling knowledge acquisition and discovery of network psychometric models. Computational psychometrics studies the computational basis of learning and measurement of traits, such as skills, knowledge, abilities, attitudes, and personality traits through methods like mathematical modelling, intelligent learning and assessment virtual systems, and computer simulation of large-scale, complex data which traditional psychometric approaches are ill-equipped to handle. Recent investigations into these hard to measure constructs include work on collaborative problem solving, teamwork, and decision making, among others. Collaborative learning is a situation in which two or more people learn or attempt to learn something together. Unlike individual learning, people engaged in collaborative learning capitalize on one another's resources and skills (asking one another for information, evaluating one another's ideas, monitoring one another's work, etc.). More specifically, collaborative learning is based on the model that knowledge can be created within a population where members actively interact by sharing experiences and take on asymmetric roles. Put differently, collaborative learning refers to methodologies and environments in which learners engage in a common task where each individual depends on and is accountable to each other. These include both face-to-face conversations and computer discussions (online forums, chat rooms, etc.). Methods for examining collaborative learning processes include conversation analysis and statistical discourse analysis.

Advantages: This mechanism helps with impressive visualization, improves educational value and overcomes language barrier.

Disadvantages: It lacks flexibility.

2.4. "Shared mental models in support of adaptive, instruction for teams using the GIFT tutoring architecture" J. Fletcher and R. Sottilare, 2017.

Under the GIFT architecture, a set of tools is available for instructors to author, deliver, and evaluate intelligent tutoring applications. An essential component of GIFT is a domain- independent pedagogical module that manages instruction based on a learner's unique information. The purpose of this pedagogical module is to tailor and induce interventions via empirically- based generic instructional strategies. The goal of this research paper is to present GIFT's engine for Macro Adaptive Pedagogy (eMAP), an algorithm in the form of a decision tree that is able to inform adaptation based on generalized characteristics associated with the learner and the targeted domains. In addition, the results of a preliminary validation study to exam the implementation of the eMAP are also included in this paper. The Generalized Intelligent Framework for Tutoring (GIFT) is an open-source architecture under development by the U.S. Army Research Laboratory and is the transition target for the work described. GIFT is an empirically-based, service-oriented framework of tools, methods and standards to make it easier to author computer-based tutoring systems (CBTS), manage instruction and assess the effect of CBTS, components and methodologies. GIFT is being developed under the Adaptive Tutoring Research Science & Technology project at the Learning in Intelligent Tutoring Environments (LITE) Laboratory, part of the U.S. Army Research Laboratory - Human Research and Engineering Directorate (ARL-HRED).

Advantages: Feature of virtual classroom with interactive facilities. It also enables large range of tutors to choose from and fool proof method to schedule of availability.

Disadvantages: Few problems which can occur through this method are lesson interruption, lack of opportunity to develop social skills and distraction from programs running on a computer.

2.5. "A tough nut to crack: Measuring collaborative problem solving" L. Lei, J. Hao, A. von Davier, P. Kyllonen, and J-D. Zapata-Rivera, Handbook of Research on technology Tools for Real-World Skill Development, 2018.

Collaborative problem solving (CPS) has been receiving increasing international attention because much of the complex work in the modern world is performed by teams. However, systematic education and training on CPS is lacking for those entering and participating in the workforce. In 2015, the Programme for International Student Assessment (PISA), a global test of educational progress, documented the low levels of proficiency in CPS. This result not only underscores a significant societal need but also presents an important opportunity for psychological scientists to develop, adopt, and implement theory and empirical research on CPS and to work with educators and policy experts to improve training in CPS. It identifies the existing theoretical frameworks and empirical research that focus on CPS and also provides examples of how recent technologies can automate analyses of CPS processes and assessments so that substantially larger data sets can be analysed and so students can receive immediate feedback on their CPS performance. Collaborative problem solving is about people working together face-to-face or in online workspaces with a focus on solving real world problems. These groups are made up of members that share a common concern, a similar passion, and/or a commitment to their work. Members are willing to ask questions, wonder, and try to understand common issues. They share expertise, experiences, tools, and methods. These groups can be assigned by instructors, or may be student regulated based on the individual student needs. The groups, or group members, may be fluid based on need, or may only occur temporarily to finish an assigned task. They may also be more permanent in nature depending on the needs of the learners. All members of the group must have some input into the decision making process and have a role in the learning process. Group members are responsible for the thinking, teaching, and monitoring of all members in the group. Group work must be coordinated among its members so that each member makes an equal contribution to the whole work. Group members must identify and build on their individual strengths so that everyone can make a significant contribution to the task. Collaborative groups require joint intellectual efforts between the members and involve social interactions to solve problems together. The knowledge shared during these interactions is acquired during communication, negotiation, and production of materials. Members actively seek information from others by asking questions. The capacity to use questions to acquire new information increases understanding and the ability to solve problems. Collaborative group work has the ability to promote critical thinking skills, problem solving skills, social skills, and self-esteem. By using collaboration and communication, members often learn from one another and construct meaningful knowledge that often leads to better learning outcomes than individual work.

Advantages: Reflects each problem at hand effectively and provides specified and detailed steps for the situation.

Disadvantages: The process might be time consuming. Since the process is a lot manual than automatic, it might be incomprehensible as well.

2.6. "Modelling behaviour dynamics using computational psychometrics within virtual worlds", Aitor Almeida, Gorka Azkune, 2018.

This paper presents a conceptual classification of the user behaviour in intelligent environments according to the different levels of granularity used to describe it. Using this classification, we also describe the algorithm developed to automatically model the inter-activity behaviour. This algorithm uses long short-term memory networks (LSTMs) to create a probabilistic model of user behaviour that can predict the next user action and detect unexpected behaviours. A virtual (unreal) world is a computer based simulated environment which may be populated by many users who can create a personal avatar, and simultaneously and independently explore the virtual world, participate in its activities and communicate with others. These avatars can be textual, two or threedimensional graphical representations, or live video avatars with auditory and touch sensations. In general, virtual worlds allow for multiple users but single player computer games, such as Skyrim, can also be considered a type of virtual world

Advantages: Predicts human behaviour efficiently based on a given dataset.

Disadvantages: Vanishing Gradient Problem occurs here. It cannot be stacked into deeper models. It is also unable to keep track of long-term dependencies.

CHAPTER 3

PROBLEM STATEMENT

3.1 Existing System

The existing system includes algorithm or manual test cases where the prediction of the human behaviour was made on statistical approach and previous actions performed by the human. Since the existing system was manual more than automated, it was bound to failures and inconsistency. In the past few years, the Artificial Intelligence (AI) and Machine Learning (ML) communities have been putting their efforts into presenting and publishing advanced methods for processing and analysing human behaviour related multi modal data.

Disadvantages:

- The existing process was not automated leading to resource and time wastage.
- Data categorization and distribution was ambiguous.
- The existing system could not identify cross-cutting capabilities or solutions.

3.2 Proposed System

This paper proposes the following three-stage method to explore and study collaborative group behaviour.

The first stage integrates and processes multimodal data obtained in a collaborative learning environment (CLE) that includes sensor input, audio, video, eye tracking, facial expressions, movement, posture, gestures, and behavioural interaction log data. The second stage performs feature extraction and cloud computation using computational psychometrics (CP) and convolutional neural network (CNN)-based deep learning for skill, pattern, and trend identification. Finally, the third stage uses the parameters measured in the previous two stages to understand and model group interactions, competencies, and collaborative behaviour at a micro-level. The third stage uses machine learning for effective assessment and visualization of group dynamics such as correctly assessing the increase in the groups' level of shared understanding of different perspectives, and ability to clarify misconceptions.

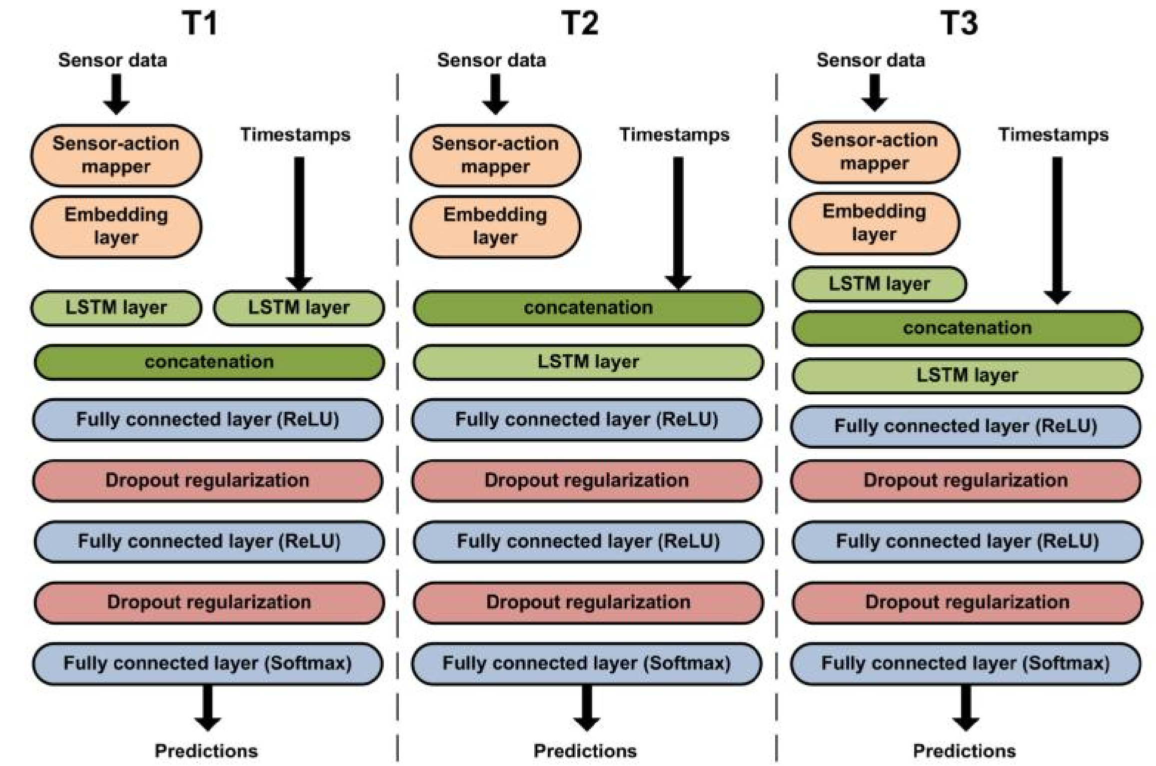

Figure 3.1: Human behavior prediction system using LSTM (Long short-term memory network).

Advantages:

- Built upon the Hadoop data analytics platform which improves the efficiency of data storage.
- Additive feature of CNN (Convolutional Neural Network) and deep learning concepts using Python or R programming makes this process a lot more efficient.
- Usage of HPC (High Performance Computing) infrastructure for massive data processing.
- Usage of machine learning for effective assessment and visualization.

CHAPTER 4

APPROACHES AND METHODS

4.1 System Architecture:

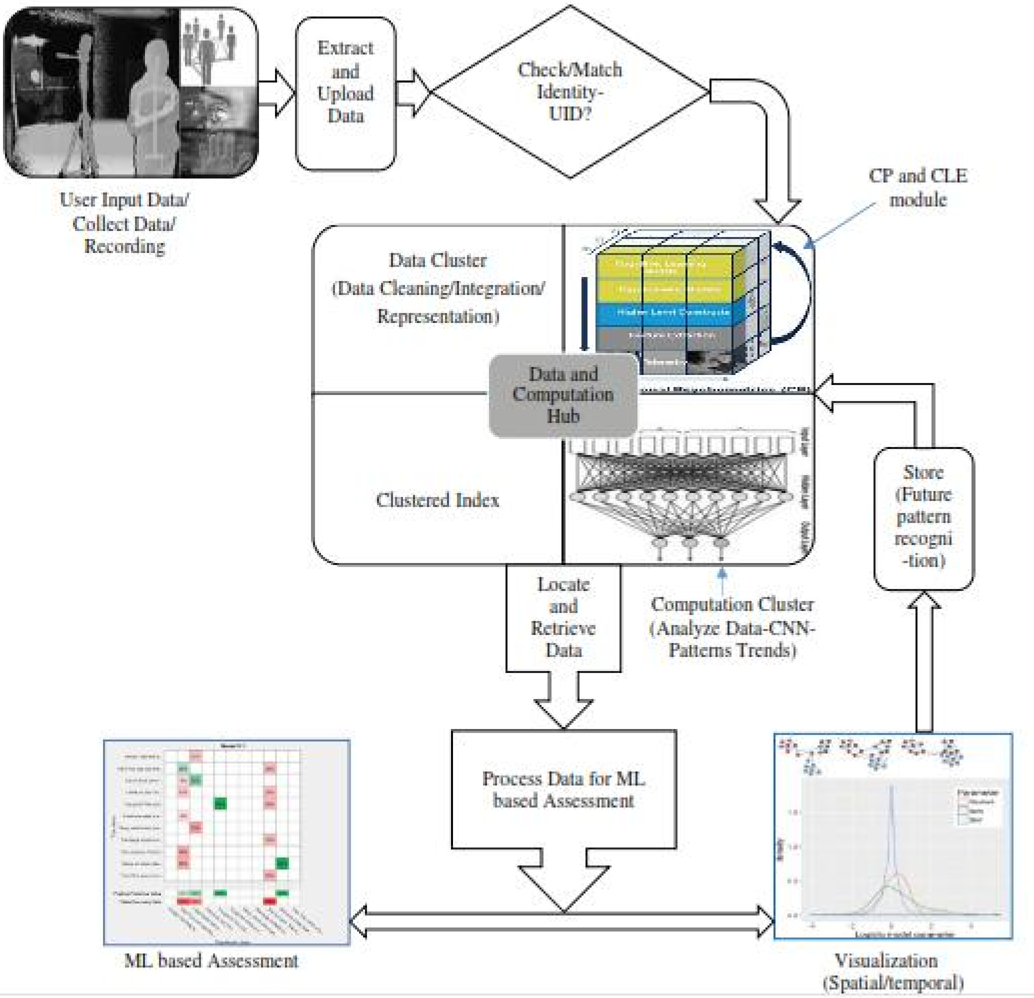

Figure 4.1: Framework for ML based data intensive computing and efficient assessment.

Fig. 4.1 shows a possible arrangement of components for machine learning (ML) based data-intensive computing and efficient assessment. Some of the components are listed here:

- Data Integration and Processing

  Establishing identities from vast volumes of CLE interaction multimodal data obtained from different sources is an essential task of data analytics and computation architectures. Large amounts of CLE multimodal interaction data that provide the identities of humans, machines, sensors, etc. collected from different sources will be processed through a set of solutions built upon the Hadoop data analytics platform.

- Massive data intensive CNN (deep learning) based cloud computing and Computational Psychometrics (CP)

  Once data has been categorized, a computation cluster will be used to analyse the data on a cloud platform to understand individual and group abilities. The programming language which will be used is Python/R to run deep learning in the cloud using ACT's enterprise learning analytics platform (LEAP).
- Effective Assessment (EA)

  The effective assessment (EA) module locates and retrieves data after the computation is finished using the CNN and CP modules illustrated in the above two modules. The CNN and CP modules carry out feature extraction, model training, and pattern identification from two player dyadic gameplay logs and behavioural data. Through effective assessment, we aim to analyse human behaviour as it relates to specific situations in the game, to detect the dynamics of group behaviour, such as shared understanding and engagement at the micro level.

4.2 Methodology:

As shown in Fig. 4.2, this study focused on collecting the following: game log data, user eye tracking, and user portrait video/audio, chat logs (conversation flow), behavioural expressions, object clicks, time in-game & between games. Our objective is to use this CPS human-human (HH) game log-data for team interaction analysis and for finding teamwork skill evidence based on the CPS construct. As illustrated in Fig. 4.2, from two player dyadic CPS game interactions, we extract four types of data files: video, eye tracking, audio, and chat/text. We then use all these source files for identity matching or for naming convention purposes. Then we perform CNN-based machine learning analysis for feature extraction, correlation, and pattern identification. We processed video files through Noldus Face reader and observer analysis which produces different behavioural markers and emotional states for dyadic gameplay (third block in Fig. 4.2).

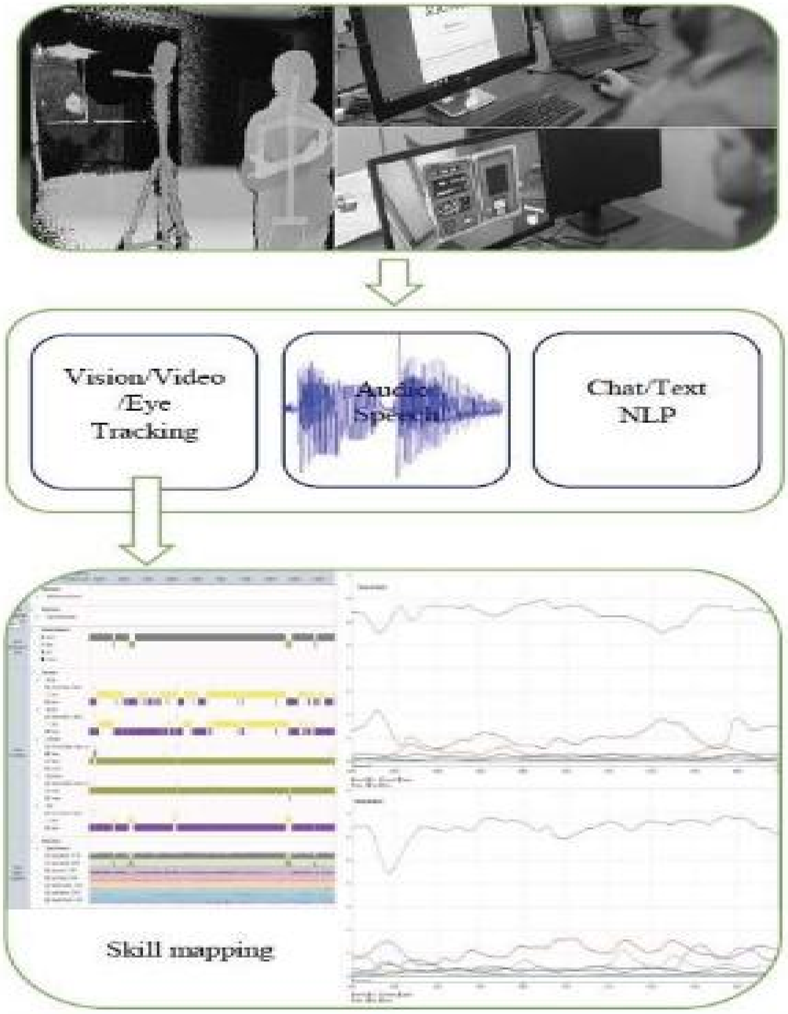

Figure 4.2: CLE multi-modal data analytics and skill mapping process.

Advantages of Machine Learning are:

- **Machine learning can easily consume unlimited amounts of data with timely analysis and assessment.**

This method helps review and adjusts your message based on recent customer interactions and behaviours. Once a model is forged from multiple data sources, it has the ability to pinpoint relevant variables. This prevents complicated integrations, while focusing only on precise and concise data feeds.

- **Machine learning algorithms tend to operate at expedited levels.**

In fact, the speed at which machine learning consumes data allows it to tap into burgeoning trends and produce real-time data and predictions. For example, machine learning can optimize and create new offers for grocery and department store customers. This means that what customers might see at 1 p.m. may be different than what they see at 2 p.m.

- **Applying machine learning to practical applications and scenarios is simply vital.**

  While predictive analytics are instrumental in saving costs and building revenue - it is equally as important to understand their impacts on real-life situations pertaining to customer acquisitions or loss.

- **Machine learning is proactive and specifically designed for "action and reaction" industries.**

  In fact, systems are able to quickly act upon the outputs of machine learning - making your marketing message more effective across the board. For example, newly obtained data may propel businesses to present new offers for specific or geo-based customers. However, data can also signify cutting back on unnecessary offers if these customers do not require them for conversion purposes.

  The latter may even be a form of learning from past behaviours. Machine learning models are able to learn from past predictions, outcomes and even mistakes. This enables them to continuously improve predictions based on new incoming and different data.

CHAPTER 5

RESULTS AND DISCUSSION

Our ML-based data-intensive computing framework has wide-ranging scalable applications including next-generation collaborative learning and assessment systems, DHS, Defence Army, Air force, Navy soldiers/Team training, and development. U.S. Army is considering learner and team-centric training, which will allow the development of mission-capable militaries, and organized teams to handle in complex situations.

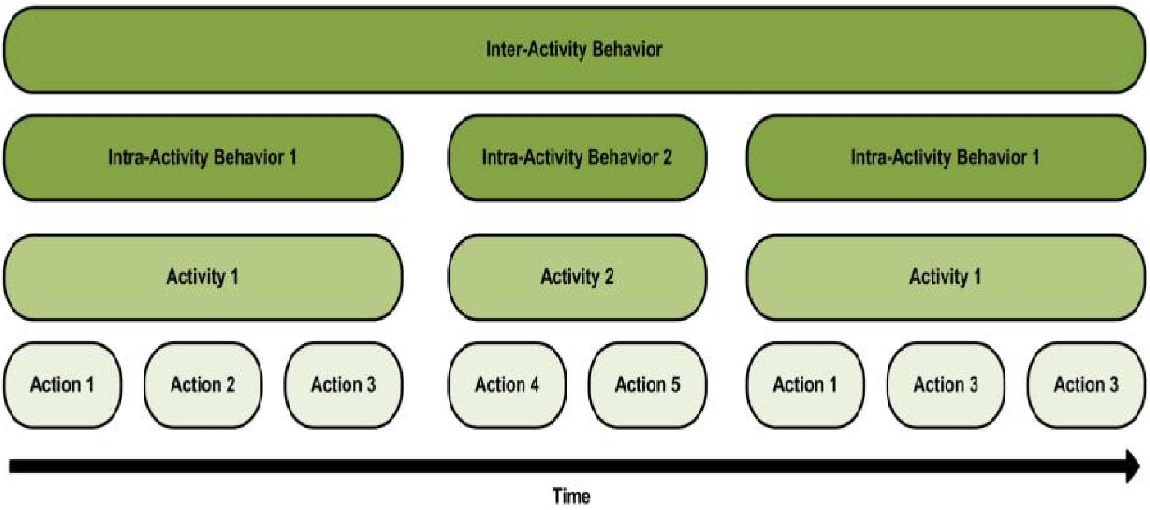

Figure 5.1: Elements of user behaviour

Figure 5.2, 5.3 and 5.4 show the results for each type of experiment. In the case of the architecture experiments (Fig. 5.2), adding more LSTM layers (A4) or smaller, densely connected layers (A5) reduced the overall accuracy of the model no matter the number of predictions. The usage of bidirectional LSTMs (A6) was also detrimental. Generally, higher dropout regularization values (A2 and A3) produced better results. Overall, the A3 configuration offered the best results across the board. As can be seen in the table, the use of embedding obtained better results when compared to the same configurations using one-hot vectors (A3 vs. A7, A2 vs. A8 and A4 vs. A9). A more semantically rich representation of the actions allows the same architecture to better model the user behaviour. The advantage of using embedding is that the training process to obtain them is completely unsupervised, not requiring a significant overheat when comparing them against other representation strategies such as one-hot vectors.

For the sequence-length experiments (Fig. 5.3), the best results were achieved using sequences with a length of 4 (S3) or 5 (A3) actions. In this specific case, the optimal sequence length was closely related with each deployment, being determined by the average action length of the activities in each scenario. This value should be adjusted in each specific case in order to achieve the best results, and it cannot be generalized.

Finally for the time experiments (Fig. 5.4), the evaluation shows that none of the proposed options to take into account the timestamps (T1, T2 and T3) improved the results. When comparing these with architecture with a similar configuration and no timestamps (A3), the results were clearly worse. Our initial intuition was that taking into account the timestamps would help to model the temporal patterns in the actions, but the results show that this information is not so relevant for the behaviour modelling task. We expect that the temporal data will be much more relevant in the activity recognition task, particularly when discriminating between activity patterns with similar actions that happen at different periods (preparing breakfast vs. preparing dinner).

Table 5.1: Architecture experiments: Accuracy for different number of predictions.

ID	acc_at_1	acc_at_2	acc_at_3	acc_at_4	acc_at_5
A1	0.4487	**0.6367**	0.7094	**0.7948**	0.8461
A2	0.4530	0.6239	**0.7222**	0.7692	0.8504
A3	**0.4744**	0.6282	0.7179	0.7905	**0.8589**
A4	0.4444	0.5940	0.6965	0.7735	0.8247
A5	0.4402	0.5982	0.7136	0.7820	0.8418
A6	0.4487	0.6068	0.7136	0.7905	0.8376
A7	0.4572	0.6153	0.7094	0.7820	0.8376
A8	0.4529	0.5811	0.7051	0.7735	0.8376
A9	0.4102	0.5940	0.7008	0.7777	0.8247

Table 5.2: **Sequence-length experiments**

ID	acc_at_1	acc_at_2	acc_at_3	acc_at_4	acc_at_5
A3	**0.4744**	0.6282	0.7179	0.7905	**0.8589**
S1	0.4553	0.5957	0.7021	0.8	0.8553
S2	0.4255	0.6255	0.7021	**0.8085**	0.8382
S3	0.4658	**0.6452**	**0.7264**	0.7948	0.8504
S4	0.4700	0.6196	0.6965	0.7692	0.8461
S5	0.4592	0.6351	0.7210	0.7896	0.8369
S6	0.4192	0.5589	0.6593	0.7554	0.8122

Table 5.3: Time experiments

ID	acc_at_1	acc_at_2	acc_at_3	acc_at_4	acc_at_5
A3	**0.4744**	**0.6282**	**0.7179**	**0.7905**	**0.8589**
T1	0.4487	0.6239	0.7094	0.7692	0.8076
T2	0.4487	0.6111	0.7008	0.7692	0.8247
T3	0.3846	0.5940	0.7051	0.7564	0.8076

A summary of the best results can be seen in Figure 5.5 which is described below.

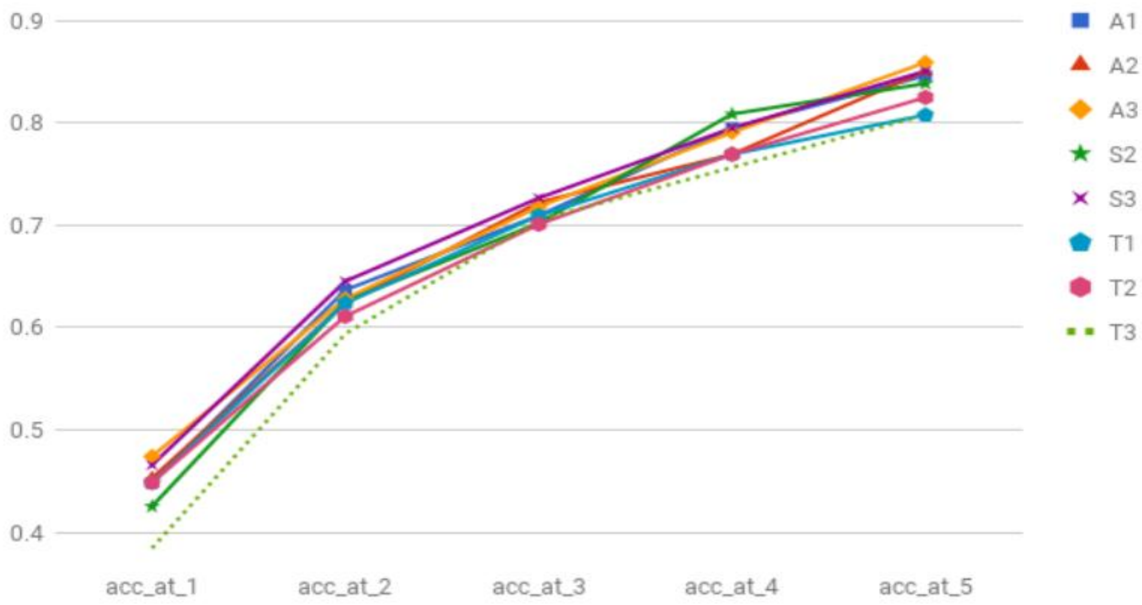

Figure 5.2: Best results in all the experiments.

CHAPTER 6

CONCLUSION

In this paper, a machine learning (ML)-based system architecture has been presented to identify evidence about teamwork skills from the behaviour, group dynamics, and interactions in the CLE. We developed a three-stage robust architecture for data intensive computing and efficient assessment of teamwork CPS skills. In this paper, we have proposed a multilevel conceptual model that describes the user behaviour using actions, activities, intra-activity behaviour and inter-activity behaviour Using this conceptual model, we have presented a deep learning architecture based on LSTMs that models inter activity behaviour. Our architecture offers a probabilistic model that allows us to predict the user's next actions and to identify anomalous user behaviours. We have evaluated several architectures, analysing how each one of them behaves for a different number of action predictions.

Future Enhancement: attempt to build text-based Natural Language Processing (NLP) / Machine Learning (ML) models to identify or classify various performances of CPS sub skills from the chat logs, audio/video interactions data collected throughout the study. Additional feature extraction that may be used during this phase will be implemented for CNN based pattern identification. The knowledge gained in developing this baseline model will represent significant guidance for proceeding phases and potential studies to follow. As future enhancement, the next step is to continue studying how the temporal and spatial information could be integrated into the architecture to improve the statistical model. We would like to explore other deep learning architectures, for example, using convolutional neural networks (CNNs) instead of LSTMs. Using CNNs, we intend to model the different action n-grams that occur in the inter-activity behaviour. This will allow us to compare different sequence modelling approaches using deep neural models. We also plan to develop architectures that will cover other aspects of the proposed multilevel conceptual model, starting with the activity recognition task. We plan to use the insights gained with the deep learning architectures proposed in this paper in order to create an activity recognition algorithm.

REFERENCES

[1]. L. Lei, J. Hao, A. von Davier, P. Kyllonen, and J-D. Zapata-Rivera. "A tough nut to crack: Measuring collaborative problem solving." Handbook of Research on Technology tools for Real-World Skill Development. IGI Global, 2016. 344 - 359. Web. 28 Aug. 2018. doi:10.4018/978-1-46669441-5.ch013

[2]. P. Cipresso, "Modeling behavior dynamics using computational psychometrics within virtual worlds," *Frontiers in Psychology, vol. 6, no.* 1725, p. 22, 6 November 2015.

[3]. J. Fletcher and R. Sottilare, "Shared mental models in support of adaptive, instruction for teams using the GIFT tutoring architecture," *International Journal of Artificial Intelligence in Education, pp. 1-21, 2017.*

[4]. A. von Davier, "Virtual and collaborative assessments: Examples, implications, and challenges for educational measurement." in Invited Talk at the Workshop on Machine Learning for Education at International Conference of Machine Learning, Lille, France, 2015.

[5]. A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *NIPS: Advances in Neural Information Processing Systems* 25, 2012.

[6]. R. Moore, B. C., R. Marciano, A. Rajasekar, and M. Wan, "Data-Intensive computing," in The Grid: Blueprint for a New Computing Infrastructure, I. Foster and C. Kesselman, Eds., New York, Morgan Kaufmann Publishers Inc, ELSEVIER, pp. 105-128, 1998

[7]. G. Grover, M. Bienkowsk, A. Tamrakar, B. Siddiquie, D. Salter, and A. Divakaran, "Multimodal analytics to study collaborative problem solving in pair programming," in *LAK '16 Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 2016.

[8]. I. Magnisalis, S. Demetriadis, and A. Karakostas, "Adaptive and intelligent systems for collaborative learning support: A review of the Field," *IEEE Transactions on Learning Technologies*, vol. 4, no. 1, pp. 520, 28 January 2011.