Towards a Neuro-Symbolic Framework for Multimodal Human-AI Interaction

Anthony James Scotte and Varuna De Silva
Institute for Digital Technologies, Loughborough University, London, U.K.

Keywords: Conversation, Emotion, Sarcasm, Neural Network, Decision Theory, Probabilistic Logic.

Abstract: Humans are not defined by a single means of communication: language, style, expression, body posture, emotion and attitude all contribute to the mix that makes communicating challenging to understand. As humans seek to strengthen partnerships with computers, these communication complexities need to be both understood and overcome. This paper seeks to explore a framework that combines neural networks with decision theory and probabilistic logic to tackle the complexities inherent within conversational communication. In combining the strength of each of these capabilities through a proof of concept, this paper demonstrates a potential framework of how different AI models may deepen its understanding of human conversation.

1 INTRODUCTION

Following decades of research, programming has evolved to a level where machines can behave in a manner that is arguably clever, for example, playing and beating human Grand Masters at the boardgame of Go (Silver et al., 2016). Further, the field of AI has sought to transcend single functions and build capabilities that partner with human beings by learning and interacting in ways familiar to their creators (Johnson, 2021). At a fundamental level, this raises questions as to how should or could computers interact with humans?

When considering how humans communicate through conversation, the words alone only convey a portion of the message (Devito, 2016). Feelings, emotions, thoughts, and beliefs, underpinned by the individual's prevailing attitude (Marchant, 2017), are inherent within both verbal (i.e., what was said, the dictum); non-verbal messages (i.e., how it was said); and what is actually meant (i.e., the implicatum) (Tabacaru, 2019). Decisions in what and how to respond in conversations require real-time inference and decision making.

Consider further the additional complexities when different languages, words and meanings, accents, cultural & behavioural norms, and colloquialisms that exist across the globe are introduced. Each additional communication aspect introduces an additional lens we as humans place across how we interpret what is being said and how we respond. The challenges for effective human to human communication are obvious and can take years to master.

This raises the not inconsiderable question of how could AI offer a solution that is able to undertake a conversation with a human that is replete with both the dictum and the implicatum and infer conclusions in near real time?

The hypothesis of this paper contended that models that identify emotion and attitude independently (in this case sarcasm as it offered an implicatum) from conversational data, could be inputs to further inference yielding a greater understanding of a conversation and thereby facilitate a decision for a better response in a possible dialogue-based solution.

The technical challenge lay in recognising that although models could perform tasks specific to particular challenges, for example emotion recognition (Chudasama et al., 2022), they were unable to extend beyond their original purpose. This required holistic end-to-end thinking that by necessity sought to integrate the capabilities of independent neural networks for low-level inference (Ghosal et al., 2020), with decision theory (Rosen, 1995) and probabilistic logic (dtai.cs.kuleuven.be, 2020) for high-level inference.

In providing a cascading framework that harnesses the strengths of low and high-level inference models, the contribution of this paper has sought to provide one possible framework that could

Scotte, A. and De Silva, V.
Towards a Neuro-Symbolic Framework for Multimodal Human-AI Interaction.
DOI: 10.5220/0011776300003411
In *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods (ICPRAM 2023)*, pages 932-939
ISBN: 978-989-758-626-2; ISSN: 2184-4313

address the nuances of language inherent within conversation and improve the chances of human to machine conversations in the future.

2 BACKGROUND

2.1 Human Considerations

Referencing the initial ‘code model’ by (Saussure, 1916), communication is seen as an exchange of information. For this to occur successfully, two conditions are thought to be required.

According to (Sperber and Wilson, 1990), the sharing of information between two human beings relies upon a shared cognitive environment which facilitates their interpretation of what they hear. Within this shared cognitive environment, all the information needed for a successful communication is available. It is therefore arguably true, that without this shared environment, the chances for a successful communication are diminished.

In a study by (Grice, 1989) and later assessed by (Davies, 2007), distinction between saying and meaning was key. According to (Grice, 1989), speakers within a conversation both express and infer ideas where speakers adhere to standard behaviours. When hearing or producing an utterance, it is assumed that the message contained all the accurate, necessary, and relevant information required. Conversely, where an utterance did not adhere to the standard behaviour, the message was not dismissed, but rather it was assumed that an appropriate meaning had to be inferred by actively seeking meaning from the context (Davies, 2007).

To complicate matters further, different attitudes, such as sarcasm and irony, can fundamentally alter what is actually said, the dictum, and what is actually meant, the implicatum (Tabacaru, 2019). Findings in one linguistic study demonstrated that 8% of utterances within conversations amongst friends contained an attitude of sarcasm (Gibbs, 2000).

To humans, understanding is not seen as sequential. The recipient “…does not first decode the logical form, then construct an explicature and select an appropriate context, and then derive a range of implicated conclusions. Comprehension is an on-line process, and hypotheses about explicatures, implicated premises, and implicated conclusions are developed in parallel against a background of expectations which may be revised or elaborated as the utterance unfolds.” (Wilson & Sperber, 2004)

2.2 AI and Conversation

Natural Language Processing (NLP), a combination of Artificial Intelligence and Linguistics, has continued to mature technologically over the last two decades and is dedicated to achieving computer comprehension of statements made by humans (Khurana et al., 2022).

As emotion is embedded within conversation, dialogue analysis has seen increasing attention from researchers (Firdaus et al., 2020) giving rise to the NLP subfield of emotion recognition in conversation (ERC).

Given the significance of the interaction between human and machine (Tu et al., 2021), interest in dialogue systems has steadily grown (Tu et al., 2022) with an increasing numbers of conversation datasets being established (Zahiri et al., 2018).

In an exploration of state-of-the-art methods used within ERC, (Chudasama et al., 2022), acknowledged that many methods still employ text-based processing and achieved robust results. However, in creating their SOTA Multi-modal Fusion Network (M2FNet) for ERC, the richness offered by multi-modal fusion exceeded associated data fusion challenges.

However, AI research has not been monopolised by emotion studies. Given the findings by (Gibbs, 2000), research has broadened into fields more focused on attitude as well (Sangwan et al. 2020) including (as examples) sarcasm detection (Ray et al., 2022) and irony detection (Reyes et al., 2012).

As imagined, sarcasm detection increases the complexity of language processing in that expressions may invert the literal meanings and therefore undermines the accuracy and robustness of NLP models (Ren et al., 2020).

Sarcasm detection therefore involves correctly identifying contextual or linguistic incongruity, which in turn requires further information, either from multiple modalities or from contextual history (Sangwan et al., 2020).

2.3 Inference from Conversation

Regardless of their increasing capabilities, one constraint on any deep learning approach is that they have been engineered to perform a single prediction task and look at the data provided through a single lens. They can predict either emotion or an attitude (for example sarcasm), but not both.

This highlights a limitation that current deep learning models can only examine one facet of a conversation but are unable to extend across all inferences and meanings. How could individual models draw inference when multiple competing explicatures are available from the same data?

Although there have been efforts to demonstrate the logical and reasoning capabilities of AI, notably using deep learning models, their results have been mixed. As examples: Deep Mind in resolving simple math questions realised 50% accuracy (only slightly better than guessing) when operations required step-by-step solving abilities (Saxton et al., 2019); and OpenAI's GPT-2 language model (Radford et al., 2019) purported to successfully generate human-like essays. However, closer inspection revealed outputs were not logically generated, but simply regurgitated examples based on the training data.

Deep learning (as a neural network) forms with other types of models a collective known as Sub-symbolic or neural AI (Yalçın, 2021). Such capabilities excel at low-level perception (Manhaeve et al., 2021a).

Symbolic AI is a sub-field of artificial intelligence, which concentrates on high-level symbolic (human-readable) representations of classical logic and assumes that logic makes machines intelligent (Yalçın, 2021). Such capabilities excel at high level reasoning (Manhaeve et al., 2021a).

The combination of deep learning and symbolic reasoning has been coined as Neuro-Symbolic AI (NeSy) (Garcez et al., 2020). NeSy's objective according to (Hamilton et al., 2022), is to combine the strengths of both symbolic and sub-symbolic approaches and address their respective weaknesses. This has seen increasing research into NeSy solutions over the last decade (Hamilton et al., 2022) where a combination of the neural and symbolic approaches is hoped to provide an opportunity to generate more robust AI solutions (Sarker et al., 2021).

One such avenue of exploration addresses implementing symbolic AI in one of the oldest (yet still extremely popular) logic programming languages, Prolog, which has its roots in first-order logic (Yalçın, 2021).

It has been further argued that high-level reasoning may be better addressed using probabilistic logic (Manhaeve et al., 2021a) and so Problog (an extension of Prolog) may offer an introduction to exploiting the possibilities of NeSy.

3 METHODOLOGY

The hypothesis of this paper contended that models that identify emotion and attitude independently (in this case sarcasm as it offered an implicatum) from conversational data could be inputs to further inference yielding greater understanding of the conversation allowing and for a better response in a possible dialogue-based solution.

In accepting that neural networks (NN's) are continuing to excel at providing low-level perception derived from data but are (at this time) less mature in providing high level inference (Manhaeve et al., 2021b), it is reasonable to propose that NN's might be coupled together with some form of inferencing capability as per the cascading model offered by (Mao et al., 2019).

In following the above reasoning, this paper sought to explore within a time-boxed, high-level proof of concept (PoC) the ability to combine the outputs from an emotion recognition neural network; a sarcasm recognition neural network; and interlocutor meta data; with decision theory and probabilistic logic from which to generate inference for decision making purposes.

A logical architecture of the PoC can be seen in Figure 1. Beyond the data sources, two individual tiers were coupled together and perform specific tasks.

3.1 Recognition Tier

The purpose of the NN's contained within the recognition tier was relatively simple in that they had to generate a probability distribution (prediction) of both an emotion, and sarcasm independently.

For the emotion recognition tier, the MELD conversation dataset and the COSMIC prediction model were used to generate a probability distribution from one of the explicit emotion categories as specified within (Ghosal et al., 2020).

For the sarcasm recognition tier the MUSTARD++ sarcastic utterance dataset and the COSMIC prediction model were used to generate a probability distribution from one of the explicit sarcasm categories as specified within (Ray et al., 2022). The MUSTARD++ dataset was chosen as it partially overlapped the MELD dataset and therefore different models could make predictions against the same data.

Both models were trained, evaluated, and employed independently of one another.

3.2 Inference Tier

The purpose of the inference tier was more complicated, twofold and remains a work in progress.

The first task was to evaluate the emotion and sarcasm predictions against one another to identify a

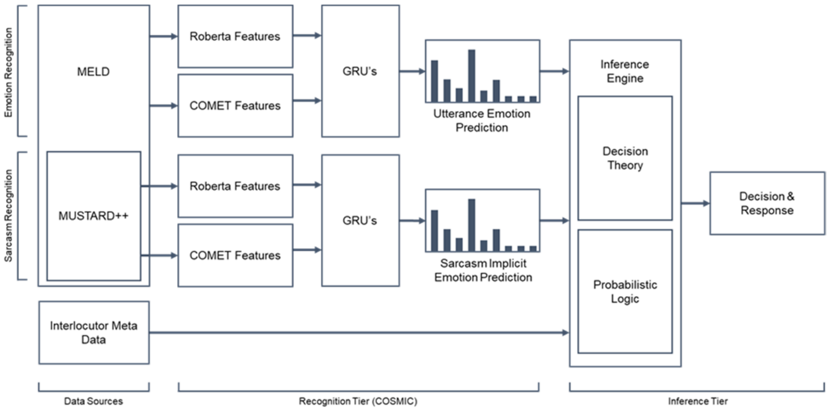


Figure 1: Proof of Concept Logical Architecture.

preferred prediction. This was attempted using Decision Theory.

The second task was to use additional interlocutor metadata with the preferred prediction as inputs and attempt to infer a deeper understanding so that a decision could be made on how to respond. This was accomplished using Probabilistic Logic.

Although NN models focus upon F1 scores to demonstrate their predictive prowess, it should be understood that a prediction by itself does not have consequences in reality until it is relied upon. Therefore, the measure of a prediction is the actual (or expected) gains (or losses) realised by those who have relied upon a prediction to decide (Rosen, 1995).

In understanding a decision problem, defined by (Rosen, 1995) as the "…choice of a course of action having real-world consequences (costs) that depend on both the action (decision) and an outcome event (true class)", it is characterised by the decision loss c_{ij} for each observed outcome i and decision j. These c_{ij} can be said to form the elements of a cost matrix C (also referred to as a decision loss matrix).

For the sake of simplicity, two outcomes and two alternatives were considered for both emotion and sarcasm although in general the number of each would arguably be higher.

Decision $j = 0$ is the most favourable when outcome $i = 0$. Nonzero c_{00} or c_{11} have been ignored as they lead to overall offsets as explained by (Rosen, 1995).

According to (Rosen, 1995), decision loss is implicit in the prediction. When considering the approach of minimising the impact of any decision, if the decision loss is known, the simplest approach would be to apply this and a prediction $\hat{p}$ into elementary decision theory and recommend the course of action $j = \hat{\jmath}$ that minimizes the expected decision loss.

$$\begin{aligned} E^i\{c_{ij}\big|\hat{p}\} &= \hat{p}c_{1j} + (1-\hat{p})c_{0j} \\ &= \begin{cases} \hat{p}c_{10} & if\ j = 1 \\ (1-\hat{p})c_{01} & if\ j = 0 \end{cases} \end{aligned} \tag{1}$$

Where: i is an observed outcome; $\hat{p}$ is the prediction of probability of outcome $i = 1$; $L(i, \hat{p})$ is the probability loss function (prediction scoring rule); j is the decision index (in some decision problem where i is relevant); $C = \{c_{ij}\}$ is the decision loss (regret or cost matrix) characterising a decision problem; t is the decision threshold = $c_{01}/(c_{01} + c_{10})$ and lies between 1 and 0; s are the stakes = $c_{01} + c_{10}$; $\hat{\jmath}$ is the decision recommendation = $I(\hat{p} > t)$; $I(inequal)$ is 1 if inequality is true, 0 otherwise; and $E^z\{g(z)\,\big|\,A\}$ is expectation over z of $g(z)$ given A.

The decision recommendation is thus given as follows:

$$\begin{aligned} \hat{\jmath} &= \begin{cases} 0 & otherwise \\ 1 & if\ (1-\hat{p})c_{01} < \hat{p}c_{10} \end{cases} \\ &\equiv I((1-\hat{p})c_{01} < \hat{p}c_{10}) \\ &= I(\hat{p} > t) \end{aligned} \tag{2}$$

It should be recognised that the decision loss is not a measure of the value of the prediction but is however given by the recommendation loss when following the recommendation implicit in $\hat{p}$ where the actual outcome is i (Rosen, 1995).

$$\begin{aligned} L_C(i, \hat{p}) &= c_{i,\hat{\jmath}} = c_{i,I(\hat{p} > t)} \\ &= \begin{cases} tsI(\hat{p} > t) & if\ i = 0 \\ (1-t)sI(\hat{p} < t) & if\ i = 1 \end{cases} \\ &= [1-i]tsI(\hat{p} > t) + i[1-t]sI(\hat{p} < t) \end{aligned} \tag{3}$$

This type of probability loss function is referred as single-decision or single-threshold since it depends only on whether $\hat{p}$ is above or below a particular t.

3.2.1 Probabilistic Logic

This part of the PoC model sought assess an individual's propensity to make jokes and the impact on the probability that the sarcastic utterance should be considered joking rather than scornful.

According to (Weitkämper, 2021), probabilistic logic programming represents a significant part of statistical relational artificial intelligence, where approaches from probability and logic are interleaved to reason from relational domains where uncertainty exists.

One such probabilistic logic program (ProbLog), which is based on First Order Logic, permits the building of programs addressing complex interactions between "…large sets of heterogenous components but also the inherent uncertainties that are present in real-life situations." (dtai.cs.kuleuven.be, 2020).

ProbLog programs themselves are comprised of a set of probabilistic facts F of the form $p :: f$ where: p is a probability; f is an atom; and $\mathcal{R}$ is a set of rules (Manhaeve et al., 2021a).

Each ground instance $f\theta$ of a probabilistic fact f, corresponds to an independent Boolean random variable with probability p when true, and probability $1 - p$ when false (Manhaeve et al., 2021a).

If all ground instances of probabilistic facts are denoted as $\mathcal{F}$ in $\mathcal{F}\Theta$, then every subset $F \subseteq \mathcal{F}\Theta$ defines a possible world $w_F = F \cup \{h\theta | \mathcal{R} \cup F \vDash h\theta$ and $h\theta$ is ground$\}$. This means that the world w_F is the canonical model of the logic program obtained by adding F to the set of rules $\mathcal{R}$ (Manhaeve et al., 2021a).

Given a possible world w_F, its probability $P(w_F)$, is given by the product of the probabilities of the truth values of the probabilistic facts (Manhaeve et al., 2021a).

$$P(w_F) = \prod_{f_i \in F} p_i \prod_{f_i \in \mathcal{F}\Theta \backslash F} (1 - p_i) \tag{4}$$

The success probability of q (also known as the probability of a ground atom q), can then be defined as the sum of the probabilities of all worlds containing q (Manhaeve et al., 2021a).

$$P(q) = \sum_{F \subseteq \mathcal{F}\Theta : q \in w_F} P(w_F) \tag{5}$$

4 RESULTS & DISCUSSION

4.1 Data

Although the MELD dataset proved a sound choice (as it continues to be used in as emotion recognition benchmark in SOTA studies), the choice of the MUSTARD++ dataset was less effective as labelled data for sarcasm modelling was limited to 601 utterances only (not conversations) across a range of American only situational comedies.

As conceded by (Ray et al., 2022), sarcasm is not easily found and required significant effort to identify and include additional conversational corpora from which to identify further data points. Despite the expansion of their search, (Ray et al., 2022), the focus remained on similar sources and were constrained by the same types of sarcasm and emotional categories.

Despite this apparent narrowness of focus, recent research by (Tabacaru ,2019) has: yielded greater insight into what sarcasm is and is not; identified the linguistic mechanisms employed in generating sarcastic utterances that are much broader than those used by (Ray et al., 2022); and broadened emotion categories which may provide greater depth to the underlying emotions inherent within sarcastic utterances.

4.2 Recognition Tier

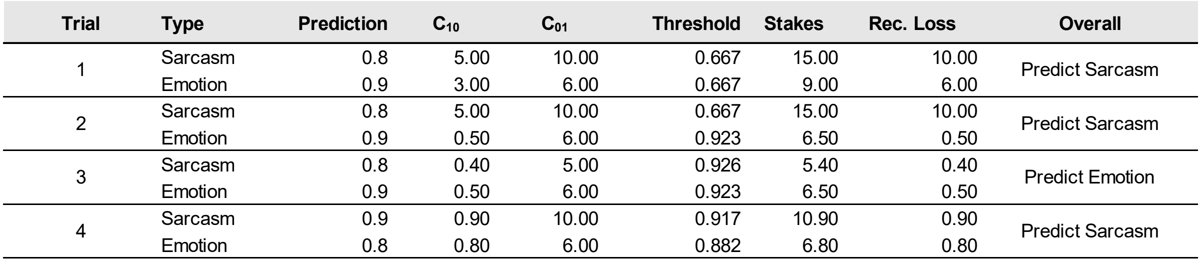
In realising the PoC, the data needs of the two prediction models were different. Although sourced from the same parent dataset, their preparation was subject to their specific needs in that the emotion predictor required conversational data whereas the sarcasm predictor required utterance data only.

In building the end-to-end capability that took conversational text and elicited predictions, the choice of COSMIC proved reasonable for the emotion recognition and results were in line with those published by (Ghosal et al., 2020).

However, the MUSTARD++ dataset, though maintaining several lines of the conversation leading up to the sarcastic utterance, did not provide emotion labels for non-sarcastic utterances. This necessitated the COSMIC model to train and validate only on those utterances that contained implicit sarcasm labels.

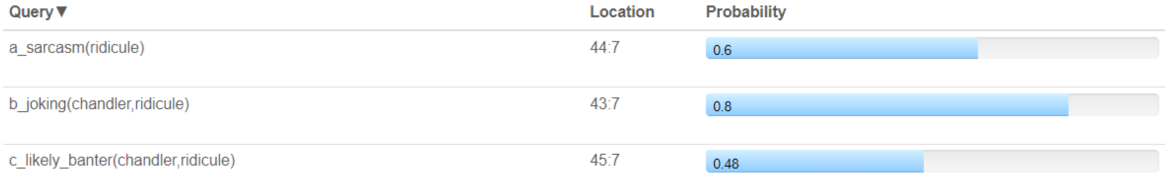
Further, the shortfall of available data constrained the model resulting in overfitting and poor results with reference to both validation and test data.

Overall, although sarcasm predictions could be generated from the model, prediction using the COSMIC model met with limited success and proxies were used within the inference tier.



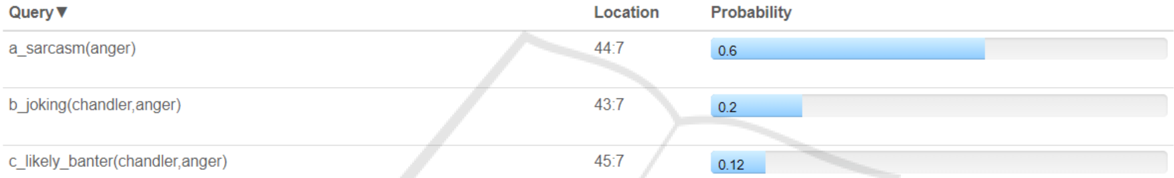
Trial	Type	Prediction	C_{10}	C_{01}	Threshold	Stakes	Rec. Loss	Overall
1	Sarcasm	0.8	5.00	10.00	0.667	15.00	10.00	Predict Sarcasm
	Emotion	0.9	3.00	6.00	0.667	9.00	6.00	
2	Sarcasm	0.8	5.00	10.00	0.667	15.00	10.00	Predict Sarcasm
	Emotion	0.9	0.50	6.00	0.923	6.50	0.50	
3	Sarcasm	0.8	0.40	5.00	0.926	5.40	0.40	Predict Emotion
	Emotion	0.9	0.50	6.00	0.923	6.50	0.50	
4	Sarcasm	0.9	0.90	10.00	0.917	10.90	0.90	Predict Sarcasm
	Emotion	0.8	0.80	6.00	0.882	6.80	0.80	

Figure 2: Sample of Decision Theory Calculations.



Query ▼	Location	Probability
a_sarcasm(ridicule)	44:7	0.6
b_joking(chandler,ridicule)	43:7	0.8
c_likely_banter(chandler,ridicule)	45:7	0.48

Figure 3: Probability of Chandler joking when sarcasm identified as Ridicule.



Query ▼	Location	Probability
a_sarcasm(anger)	44:7	0.6
b_joking(chandler,anger)	43:7	0.2
c_likely_banter(chandler,anger)	45:7	0.12

Figure 4: Probability of Chandler joking when sarcasm identified as Anger.

The limited success of the COSMIC model to successfully train on the single-modal input as opposed training on multi-modal data as realised by (Ray et al., 2022), suggests that approaches that use single modal only for sarcasm identification, or indeed any attitudinal identification, remain less effective. This is borne out by (Tabacaru, 2019) who suggests that facial features and vocal prosody are key considerations in being able to identify sarcasm.

4.3 Inference Tier

The two main findings from the application of Decision Theory (Rosen, 1995) were that: the relationship between the prediction and the threshold directly impacted the choice of the recommendation loss; and having established the decision recommendation loss, the choice of prediction was down to avoiding the biggest loss. Sample results can be seen in Figure 2. Overall, the theory proffered by (Rosen, 1995) does provide a framework for differentiating and choosing between competing predictions, so long as the costs of all decisions can be sufficiently quantified.

As this study was an exercise seeking to explore and prove a concept, certain assumptions were made in the number of possible alternative decisions that could occur for each outcome; and determining the values to be attributed to the decision costs applied within the cost matrix.

In reality, the alternatives and costs arguably may be able to be provided by experts within the field. However, where this is not possible and these remain unknown, (Rosen, 1995) proffers that the recommendation loss can be plotted as a function of the unknown threshold for a given prediction and outcome. In summing or averaging such curves over a data set consisting of many prediction/outcome pairs, the performance on the data can be characterised. This (Rosen, 1995) refers to as the recommendation loss characteristic (RLC) curve and will provide a basis over time from which a threshold can be derived and updated.

In applying ProbLog to possible treatments of additional interlocutor metadata such as propensity to banter when the sarcasm prediction was ridicule, ProbLog does provide a successful coding framework. Examples can be seen in Figures 3 & 4.

However, in extracting and making separate the logic that determined if the sarcastic utterance of the individual was banter based on the individual's predilection for making jokes, the question was raised as to where the logic should reside.

According to (Manhaeve et al., 2021b) there is a prevailing idea that employs logic as a constraint within a deep model. This is achieved by extending a deep model with a regularisation term, drawn from desired logical properties that encourages the model to imitate logical reasoning or suffer a penalty if not obeyed.

In using a regulariser encouraging a model to satisfy constraints, the logic is encoded into the parameters (either weights or embeddings). Arguably, the constraints are soft, and there is no guarantee that they all will be satisfied or make predictions coherent with the logic they were trained on (Xu et al., 2018).

Regardless, (Manhaeve et al., 2021b) contends that the underlying importance of being able to fully recover the logic is the primary objective despite the location of the logic.

In the case of the PoC and this study, additional lenses across the conversational data may require additional and deeper logics to be added over time. Implementing logical constraint via a regulariser across all models may arguably not prove effective considering there is no guarantee they will all work as intended and thereby result in inconsistent results when compared.

5 CONCLUSIONS

The contribution of this paper has been to demonstrate the nuances of language and associated modelling challenges, and identify one possible framework that, if explored further, might improve human to machine conversations in the future.

Although the solution and evaluation remained high-level due to the time-boxed nature of the study, conclusions were able to be drawn and provide insights for future versions.

What data exists at present for emotion and attitude is not yet of sufficient quality or quantity. For such a solution to become more robust, a cross-disciplinary, multi-modal research exercise with linguists should be undertaken to fully understand the latest research, and elicitation strategies to gather an appropriately qualified and robust dataset that serves a diverse set of requirements.

It can be further argued that in bringing multiple models together to contribute to one holistic purpose, the issue of aligning models across multiple modalities to evaluate the same data or use similar algorithm techniques to ensure consistent training and prediction, does require careful consideration, and an end-to-end holistic design mindset.

The decision-making process illustrated choosing between competing predictions can be made with an understanding of the associated costs incurred when relying upon a given prediction. Although this framework did rely upon a simplistic case and cost assumptions, decision recommendation loss theory does offer a working solution and provides an avenue forward for further exploration under more stringent and realistic conditions.

The application of first order logic (through ProbLog) to add further inference to a decision made earlier in the process, demonstrates that the cascading approach adopted in this study is worthy of further consideration. Appreciating that the inference undertaken in this case was simplistic in nature does not erode the underlying proposition but proved its validity and provides an avenue to increase the maturity of this capability further.

Despite the solution being a time boxed PoC, the study has explored the stated hypothesis; has identified challenges to be overcome; and laid foundations for one possible solution. However, it remains a work in progress.

Next steps for this avenue of research are to deepen the exploration and realisation of the preliminary recommendations highlighted above, create a more substantive version of the solution that incorporates findings and more fully demonstrates an end-to-end capability, and implement a framework that more comprehensively evaluates and provides deeper results for consideration.

REFERENCES

Chudasama, V., Kar, P., Gudmalwar, A., Shah, N., Wasnik, P. and Onoe, N. (2022). M2FNet: Multi-modal fusion network for emotion recognition in conversation. [online] doi:10.48550/ARXIV.2206.02187.

Davies, Bethan L. (2007). Grice's Cooperative Principle: Meaning and rationality. Journal of Pragmatics 39. 2308—2331.

Devito, J.A. (2016). The interpersonal communication book. 14th ed. Boston: Pearson.

dtai.cs.kuleuven.be. (2020). Introduction. — ProbLog: Probabilistic Programming. [online] Available at: https://dtai.cs.kuleuven.be/problog/.

Firdaus, M., Chauhan, H., Ekbal, A. and Bhattacharyya, P. (2020). EmoSen: Generating Sentiment and Emotion Controlled Responses in a Multimodal Dialogue System. IEEE Transactions on Affective Computing, pp.1–1. doi:10.1109/taffc.2020.3015491.

Garcez, A. d'Avila and Lamb, L.C. (2020). Neurosymbolic AI: The 3rd wave. [online] doi:10.48550/ARXIV. 2012.05876.

Ghosal, D., Majumder, N., Gelbukh, A., Mihalcea, R. and Poria, S. (2020). COSMIC: COmmonSense knowledge for eMotion identification in conversations. [online] doi:10.48550/ARXIV.2010.02795.

Gibbs, R. (2000). Irony in Talk Among Friends. Metaphor and Symbol, 15(1), pp.5–27. doi:10.1207/s15327868ms151&2_2.

Grice, H.P. (1989). Studies in the way of words. Cambridge, Mass.: Harvard Univ. Press, Ca.

Hamilton, K., Nayak, A., Bozic, B. and Longo, L. (2022). Is neuro-symbolic AI meeting its promise in natural language processing? A structured review.

Johnson, J. (2021). AI/Human Augmentation: How AI & Humans Can Work Together. [online] BMC Blogs.

Khurana, D., Koli, A., Khatter, K. and Singh, S. (2022). Natural language processing: state of the art, current trends and challenges. Multimedia Tools and Applications. doi:10.1007/s11042-022-13428-4.

Manhaeve, R., Dumančić, S., Kimmig, A., Demeester, T. and De Raedt, L. (2021a). Neural probabilistic logic programming in DeepProbLog. Artificial Intelligence, 298, p.103504. doi:10.1016/j.artint.2021.103504.

Manhaeve, R., Marra, G., Demeester, T., Dumancic, S., Kimmig, A. and De Raedt, Luc (2021b). Chapter 7. Neuro-symbolic AI = neural + logical + probabilistic AI. doi:10.3233/FAIA210354.

Mao, J., Gan, C., Kohli, P., Tenenbaum, J.B. and Wu, J. (2019). The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision. [online] doi:10.48550/ARXIV.1904.12584.

Marchant, J. (2017). » Attitude. [online] www.emotionalintelligenceatwork.com. Link.

Mehrabian A., Communicating without words, Psychol. Today. (1968) 53–55.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D. and Sutskever, I. (2019). Language models are unsupervised multitask learners. OpenAI blog, 1, p.9.

Ray, A., Mishra, S., Nunna, A. and Bhattacharyya, P. (2022). A multimodal corpus for emotion recognition in sarcasm. [online] doi:10.48550/ARXIV.2206.02119.

Ren, L., Xu, B., Lin, H., Liu, X. and Yang, L. (2020). Sarcasm Detection with Sentiment Semantics Enhanced Multi-level Memory Network. Neurocomputing, 401, pp.320–326. doi:10.1016/j.neucom.2020.03.081.

Reyes, A., Rosso, P. and Veale, T. (2012). A multidimensional approach for detecting irony in Twitter. Language Resources and Evaluation, 47(1), pp.239–268. doi:10.1007/s10579-012-9196-x.

Rosen, D. (1995). How good were those probability predictions? The expected recommendation loss (erl) scoring rule. Maximum Entropy and Bayesian Methods.

Sarker, M.K., Zhou, L., Eberhart, A. and Hitzler, P. (2021). Neuro-symbolic artificial intelligence: Current trends. [online] doi:10.48550/ARXIV.2105.05330.

Sangwan S., Akhtar M. S., Behera P. and Ekbal A., (2020), I didn't mean what I wrote! Exploring Multimodality for Sarcasm Detection, International Joint Conference on Neural Networks (IJCNN), pp. 1-8, doi: 10.1109/IJCNN48605.2020.9206905.

Saussure, Ferdinand de. (1916). Cours de linguistique générale. Paris: Payot

Saxton, D., Grefenstette, E., Hill, F. and Kohli, P. (2019). Analysing mathematical reasoning abilities of neural models. [online] doi:10.48550/ARXIV.1904.01557.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T. and Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), pp.484–489. doi:10.1038/nature16961.

Sperber, D. and Wilson, D. (1990). Relevance communication and cognition. Oxford Blackwell.

Tabacaru, S. (2019). A Multimodal Study of Sarcasm in Interactional Humor. Berlin: De Gruyter Mouton.

Tu, G., Wen, J., Liu, H., Chen, S., Zheng, L. and Jiang, D. (2021). Exploration meets exploitation: Multitask learning for emotion recognition based on discrete and dimensional models. Knowledge-Based Systems, p.107598. doi:10.1016/j.knosys.2021.107598.

Tu, G., Wen, J., Liu, C., Jiang, D. and Cambria, E. (2022). Context- and Sentiment-Aware Networks for Emotion Recognition in Conversation. IEEE Transactions on Artificial Intelligence, pp.1–1. doi:10.1109/tai.2022.3149234.

Weitkämper, F. (2021). An asymptotic analysis of probabilistic logic programming, with implications for expressing projective families of distributions. [online] doi:10.48550/ARXIV.2102.08777.

Xu, J., Zhang, Z., Friedman, T., Liang, Y. and Broeck, G., 2018, July. A semantic loss function for deep learning with symbolic knowledge. In International conference on machine learning (pp. 5502-5511). PMLR.

Yalçın, O.G. (2021). Symbolic vs. Subsymbolic AI Paradigms for AI Explainability. [online] Medium. Link.

Zahiri S. M., Choi J.D., (2018), Emotion Detection on TV Show Transcripts with Sequence-based Convolutional Neural Networks https://doi.org/10.48550/arXiv.1708.04299.