CHARET: Character-centered Approach to Emotion Tracking in Stories

Extended Abstract

Diogo S. Carvalho
INESC-ID & Instituto Superior Técnico, Universidade de Lisboa
diogo.s.carvalho@tecnico.ulisboa.pt

Joana Campos
INESC-ID
joana.campos@inesc-id.pt

Manuel Guimarães
INESC-ID & Instituto Superior Técnico, Universidade de Lisboa
manuel.m.guimaraes@tecnico.ulisboa.pt

Ana Antunes
INESC-ID & Instituto Superior Técnico, Universidade de Lisboa
ana.j.antunes@tecnico.ulisboa.pt

João Dias
INESC-ID & Faculdade de Ciências e Tecnologia, Universidade do Algarve & CCMAR
jmdias@ualg.pt

Pedro A. Santos
INESC-ID & Instituto Superior Técnico, Universidade de Lisboa
pedro.santos@tecnico.ulisboa.pt

ABSTRACT

Autonomous agents that can engage in social interactions with a human is the ultimate goal of a myriad of applications. A key challenge in the design of these applications is to define the social behavior of the agent, which requires extensive content creation. In this research, we explore how we can leverage current state-of-the-art tools to make inferences about the emotional state of a character in a story as events unfold, in a coherent way. We propose a character role-labelling approach to emotion tracking that accounts for the semantics of emotions. We show that, by identifying actors and objects of events and considering the emotional state of the characters, we can achieve better performance in this task when compared to end-to-end approaches.

KEYWORDS

Socially Intelligent Agents; Semantic Role Labeling; Commonsense Inference; Emotion Classification

ACM Reference Format:

Diogo S. Carvalho, Joana Campos, Manuel Guimarães, Ana Antunes, João Dias, and Pedro A. Santos. 2021. CHARET: Character-centered Approach to Emotion Tracking in Stories: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

1 INTRODUCTION

Building computer artifacts that engage in the complex social dance, particularly in open-ended domains, is a compelling prospect that has attracted many researchers. Agent-based attempts to simulate individual cognitive and affective processes [5, 7, 9, 12], model how individual traits, goals, beliefs and actions interact to produce intelligent and emotionally plausible behaviour, in any scenario that the user can imagine. Yet, it is up to the author of a scenario to manually describe them for each character and guarantee plot adaptability and consistency as events unfold. While this can be manageable in narrow domains of application, scenario complexity can escalate rapidly, reducing the power of such architectures by relying heavily on the accessibility of the tool [10] and foremost the authors' creativity and ability to anticipate all interaction paths.

In the SLICE Project[1], the ultimate goal is to create agents that can effectively act socially on the users' actions, without relying on hand-generated content only. We focus on a central element of social interactions – *emotions*. We are interested in detecting and understanding the user's (or a character) emotional state, as interactions evolve in ways that were not accounted for. Works for detecting sentiment and emotions in text take different approaches. Some tasks explore the positive or negative orientation of words [14], while other intend to infer a driving sentiment or opinion in a whole document [1]. These works focus on extracting an emotional label (or overall sentiment) from a set of words and do not consider the emotional state of the different entities in the text.

In this paper, the emotion classification task is modeled as a character role-labeling problem, because we are interested in who felt the emotion and why. Our character-centered approach diverges from common approaches to sentiment analysis in text that attempt to infer an emotion from a set of words, ignoring the semantics and subjectiveness of emotions. We find that by identifying the characters' roles in a story and keeping track of their emotional state, we can perform better than end-to-end approaches in the task of emotion tracking in stories.

2 CHARET

CHARET, a **CH**aracter-centered **A**pproach to **E**motion **T**racking in Stories, recognizes and tracks emotions in stories by identifying stimuli and their objects, aided by semantic role-labelling [8]. Given a story S, consisting of n events described in natural language $(s_1, \dots, s_n)$ and a set of m characters $C = \{c_1, \dots, c_m\}$, we assume that the emotional episode of an event-character pair (s_t, c_i) is a set of emotional reactions Y_{s_t,c_i}. The set Y_{s_t,c_i} is a subset of the set of all possible N emotions $P = \{emotion_1, \dots, emotion_N\}$. We highlight that the empty set, corresponding to *no emotion*, and the whole set, corresponding to *all emotions*, are allowed. Possible choices of P include Ekman's six basic emotions [6] or Plutchik's eight basic emotions [11]. Provided S and C, we set ourselves on

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.
© 2021 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

[1] https://gaips.inesc-id.pt/slice/

the task of tracking the emotions in Y_{s_t,c_i} with the support of commonsense inference tools. For each event-character pair (s_t, c_i) and for each emotion y in P, a function $f_y : S \times C \rightarrow \{0, 1\}$ predicts whether y is an emotional reaction of the state-character pair (s_t, c_i) based on a $score_{s_t,c_i,y} \in [0, 1]$. If the score is sufficiently high, we classify y as an emotional reaction. Once we follow the approach for every emotion y from P, we are left with the predicted set of emotional reactions $\hat{Y}_{s_t,c_i}$. The function f encompasses a pipeline with three components: *a*) character role-labeling, *b*) commonsense inference, and *c*) emotion classification. It consists in identifying stimuli (events) and their objects, and then use a language model – *COMET* [3] – to make inferences about life events and identify the triggered emotions,from the perspective of the character. Figure 1 shows a diagram of our approach over the first two examples of a story from the StoryCommonsense[2] dataset, *Hot Coffee*. For more details of our approach refer to the preprint [4].

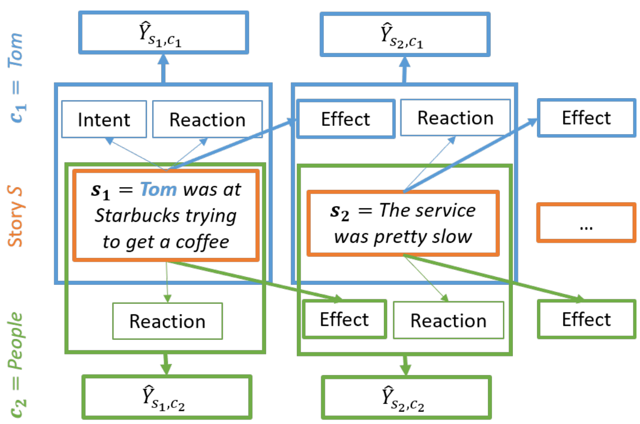

Figure 1: Diagram of our approach. *Effects* from the first event are used to classify the emotions of the second.

3 EXPERIMENTAL EVALUATION

Our algorithm aims to correctly label the emotional reactions of characters in a story. For evaluating the effectiveness of our approach we used the The StoryCommonsense dataset [13]. It consists in short commonsense stories, with five natural language events each, and is annotated with the mental states of the characters - emotional reactions. The task in which CHARET is evaluated consists in labeling the emotional reactions of the characters in each story of the StoryCommonsense dataset. Particularly, we label each story event-character pair with a subset of the eight Plutchik basic emotions used to annotate the dataset - *surprise*, *disgust*, *sadness*, *joy*, *anger*, *fear*, *trust*, *anticipation*.

3.1 Results

Table 1 shows the main results obtained on the task, compared with baselines [2]. Our approach appears in the table with the name CHARET; the approaches from Bosselut et al. with the names COMET - Direct and COMET - DynaGen. The best results of each learning setting - zero-shot and few-shot - are bolden. We also include a supervised setting to illustrate the extensibility of the apporoach. The results indicate the benefits of our approach both in the zero-shot and few-shot scenario. Additionally, we notice that the performance increases as the level of task-specific knowledge increases - from zero-shot to few-shot; from few-shot to supervised.

Table 1: Performance of ours and previous approaches to the StoryCommonsense emotion classification task.

Model	Precision	Recall	F1
Zero-shot			
Random	20.6	20.8	20.7
COMET - Direct	37.4	36.9	37.2
COMET - DynaGen	**38.9**	39.3	39.1
CHARET	31.1	**77.4**	**44.3**
Few-shot			
COMET - DynaGen	31.2	65.1	42.2
CHARET	**39.4**	**81.5**	**53.1**
Supervised			
CHARET	46.4	82.7	59.5

Table 2: Performance of character role-labeling.

Precision	Recall	F1
89.0	63.5	74.1

3.1.1 *Errors introduced by the tools used for caracter role-labelling.*

We evaluate how much error is introduced by the character role-labelling step, consisting of co-reference solution and semantic role-labling. Table 2 shows the results. The results show that our character-role labeling step is precise to some extent, as 89% of the character we classify as actors of events are correctly classified.

4 DISCUSSION AND CONCLUSION

CHARET outperforms previous works that do not account for the semantics and subjectiveness of emotions when inferring the emotional state of characters. The results obtained for the few-shot setting - requiring a small amount of training - reached an F1 score of 53.1, a 25.8 % increase over the baseline COMET - DynaGen. These results are promising, as they suggest that a layered approach to emotion tracking and prediction can yield better results than end-to-end approaches. Our work points out that an approach that combines data and semantic analysis should be considered in order to create more believable interactions between IVAs and humans. At the same time, in the real-world applications, we should expect to have more noisy data points, with a poorer structure, which could result in poorer results. Predicate and argument extraction, the backbone of our approach, may be more challenging under those conditions as shown in open-information scenarios.

ACKNOWLEDGMENTS

This work was partially supported by national funds through Fundação para a Ciência e Tecnologia under project SLICE with reference PTDC/CCI-COM/30787/2017, University of Lisbon and Instituto Superior Técnico and INESC-ID multi annual funding with reference UIDB/50021/2020.

[2]https://uwnlp.github.io/storycommonsense/

REFERENCES

[1] Rushlene Kaur Bakshi, Navneet Kaur, Ravneet Kaur, and Gurpreet Kaur. 2016. Opinion mining and sentiment analysis. In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*. IEEE, 452–455.

[2] Antoine Bosselut, Ronan Le Bras, , and Yejin Choi. 2021. Dynamic Neuro-Symbolic Knowledge Graph Construction for Zero-shot Commonsense Question Answering. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*.

[3] Antoine Bosselut, Hannah Rashkin, Maarten Sap, Chaitanya Malaviya, Asli Celikyilmaz, and Yejin Choi. 2019. COMET: Commonsense transformers for automatic knowledge graph construction. *arXiv preprint arXiv:1906.05317* (2019).

[4] Diogo S. Carvalho, Joana Campos, Manuel Guimarães, Ana Antunes, João Dias, and Pedro A. Santos. 2021. CHARET: Character-centered approach to emotion tracking in stories. *arXiv preprint arXiv* (2021).

[5] Joao Dias and Ana Paiva. 2005. Feeling and reasoning: A computational model for emotional characters. In *Portuguese conference on artificial intelligence*. Springer, 127–140.

[6] Paul Ekman. 1992. Are there basic emotions? (1992).

[7] Arno Hartholt, David Traum, Stacy C. Marsella, Ari Shapiro, Giota Stratou, Anton Leuski, Louis-Philippe Morency, and Jonathan Gratch. 2013. All Together Now: Introducing the Virtual Human Toolkit. In *13th International Conference on Intelligent Virtual Agents*. Edinburgh, UK.

[8] Daniel Jurafsky and James H. Martin. 2009. *Speech and Language Processing (2nd Edition)*. Prentice-Hall, Inc., USA.

[9] Stacy C Marsella and Jonathan Gratch. 2009. EMA: A process model of appraisal dynamics. *Cognitive Systems Research* 10, 1 (2009), 70–90.

[10] Samuel Mascarenhas, Manuel Guimarães, Rui Prada, João Dias, Pedro A Santos, Kam Star, Ben Hirsh, Ellis Spice, and Rob Kommeren. 2018. A virtual agent toolkit for serious games developers. In *2018 IEEE Conference on Computational Intelligence and Games (CIG)*. IEEE, 1–7.

[11] Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. In *Theories of emotion*. Elsevier, 3–33.

[12] Alexandru Popescu, Joost Broekens, and Maarten Van Someren. 2014. Gamygdala: An emotion engine for games. *IEEE Transactions on Affective Computing* 5, 1 (2014), 32–44.

[13] Hannah Rashkin, Antoine Bosselut, Maarten Sap, Kevin Knight, and Yejin Choi. 2018. Modeling naive psychology of characters in simple commonsense stories. *arXiv preprint arXiv:1805.06533* (2018).

[14] Theresa Wilson, Janyce Wiebe, and Paul Hoffmann. 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In *Proceedings of human language technology conference and conference on empirical methods in natural language processing*. 347–354.