Time-aware Predictions of Moments of Change in Longitudinal User Posts on Social Media

Anthony Hills[1], Adam Tsakalidis[1,2], and Maria Liakata[1,2,3]

[1] Queen Mary University of London
[2] The Alan Turing Institute
[3] University of Warwick
{a.r.hills,a.tsakalidis,m.liakata}@qmul.ac.uk

Abstract. Capturing changes in an individual's language is an important aspect of personalised mental health monitoring. A key component is modelling the influence of time, as contextual information both in the recent or distant past/future carries varying semantic weight. We capture and contrast this information by identifying neural, time-sensitive, bi-directional representations of individuals – modelling time-intervals in their social-media posts inspired by the Hawkes process. We demonstrate that our approach helps identify whether an individual's mood is changing drastically, or smoothly on two social media datasets – yielding superior performance compared to time-insensitive baselines and outperforming the state-of-the-art on the CLPsych 2022 shared task.

Keywords: Social Media · Mental Health · Longitudinal Modelling

1 Introduction

Mental health has rapidly become one of the most prevalent public health problems worldwide. A recent large-scale longitudinal study [17] on identifying changes in mental health showed that the prevalence of likely mental health problems in the UK increased from 24.3% between 2017-19 to 37.8% in April 2020 during the COVID-19 pandemic, claiming large health and economic costs and highlighting the timeliness and need for developing scalable tools for monitoring changes in mental health in an automated, real-time manner.

Social media provide a rich resource for addressing this challenge. However, most related work based on longitudinal social media data typically ignore the time-varying nature of an individual's mental state and instead make aggregate user- [3, 6, 18, 30] or post-level predictions – e.g. identifying posts with suicidal ideation indicators [13, 29, 35] or mental health symptoms [12, 24, 27].

Task Definition. This work focuses on **predicting moments of change in mood** (MoCs) in individuals on the basis of textual content shared online. Following the definitions by [38], we work with *user timelines* (sequences of

https://github.com/Maria-Liakata-NLP-Group/time-aware-predictions-of-mocs chronologically ordered individual user posts), aiming at classifying each post as a: (1) *switch* – sudden, abrupt change in mood, from positive to negative, or vice versa; (2) *escalation* – gradual change, where the user's mood progressively becomes more positive/negative; or (3) *none* – user's baseline mood, no MoC. The task formed the basis of the latest CLPsych shared task [37], where neural networks, such as LSTMs, were employed by nearly all teams [1, 4, 7, 9, 14, 28]. While such approaches help in modelling the sequence dynamics in posts, they lack the ability to account for the significant changes and heterogeneity in time-intervals between posts.

Temporal point processes (TPPs) [15, 16] are designed for modelling variable-length, asynchronous event sequences spaced irregularly in continuous-time (e.g. posts on social media). The self-exciting **Hawkes process** [25] is a popular TPP that has been frequently applied to social media data [31]. It models the probability of an event occurring, where recent events spike the probability of future events occurring, followed by a (typically exponential) decay, making the probability of future events less likely in the absence of recent ones. Neural TPPs (NTTPs) [34] leverage neural networks (typically RNNs [20]) to learn a highly complex, flexible representation of an event history, which is then used to parameterize the conditional distribution of the probability of a next event occurring.

In this paper, rather than predicting the probability of post events occurring, we instead aim to enrich learned dynamics of individuals by taking advantage of the parameterizations for modelling time in the Hawkes process. We thus model BiLSTM hidden states over post embeddings of a given user to learn the basic sequence dynamics of a user's linguistic posting context, and aggregate these with a transformation similar to a Hawkes process – to combine the learned dynamics of the sequence information captured by LSTMs with the time-sensitive information in the representations, modelled with exponential decay. Specifically we make the following contributions:

- We propose a time-aware approach for modelling textual posts of individuals, by transforming their respective LSTM hidden states over previous/future posts with self-excitation and exponential decay that varies with time.
- We extend the time-aware approach to the bi-directional setting and work on two social media datasets, showcasing that this combination outperforms non-time-aware baselines, and all teams from the CLPsych 2022 shared task.
- We demonstrate the effectiveness of our approach, in an ablation study investigating (1) time-aware features and (2) bi-directionality.

2 Method

Notation. A user's u timeline $T^{(u)}$ consists of chronologically ordered posts $p^{(u,i)} = \{v^{(u,i)}, t^{(u,i)}, y^{(u,i)}\}$ $(0 \leq i < |T^{(u)}|)$, where the i^{th} post is represented by its post-level embeddings $v^{(u,i)}$, associated posting timestamp $t^{(u,i)}$ and ground-truth label $y^{(u,i)} \in \{S, E, O\}$ (switch, escalation or none respectively). Our aim is to predict each label $\hat{y}^{(u',i)}$ in a test user's u' timeline $T^{(u')}$, given the sequence of $\{v^{(u',i)}, t^{(u',i)}\}$.

2.1 Model

Here we outline our model for incorporating temporally contextual information for identifying MoCs in a timeline. We model historical, time-sensitive information about the user (**HEAT**) and extend our model to operate over trainable hidden representations (**Modelling sequence dynamics**) as well as contextual future information (**Bi-directionality**).

Compared to prior work for this task [37, 38], we note that we are first to approach the problem with a sequence-based, bi-directional, and time-sensitive approach that considers the time-stamps of historical and future posts to predict moments of change in mood. Most prior approaches fail to model the influence of time [2, 5, 8, 10, 11, 22, 23, 28], neglect the importance of bi-directionality [2, 8, 10, 11, 14, 28] or fail to model sequence dynamics [11, 28] for predicting moments of change.

HEAT. To aggregate historical posts of a user into a temporally-informed embedding, [33] proposed to model the influence of time on an individual's historical post representations $v^{(i)}$ in a timeline via the Historical Emotional AggregaTion (HEAT), comparing each post at index i, with each historical post in the timeline at index j:

$$v_{HEAT}^{(i)} = \sum_{j:\Delta\tau_j>0} v^{(j)} + \epsilon e^{-\beta \Delta\tau_j} max(v^{(j)}, 0), \tag{1}$$

where $\Delta\tau_j = t^{(i)} - t^{(j)}$ (measured in days), and ϵ and β are fixed hyper-parameters reflecting the amount of self-excitation and exponential time-decay to apply to each post respectively when building an aggregate representation at each time-step. As such, HEAT encodes the dynamics of historical post representations in a time-aware manner.

Modelling sequence dynamics. [32] model raw post representations $v^{(i)}$ via Eq. 1. By contrast we model BiLSTM hidden state representations $h^{(i)}$ of posts, to better capture sequence dynamics. We thus apply Eq. 1 by substituting v with h at each timestep to model aggregate representations of u's historical posts.

Bi-directionality. Motivated by the notion that changes are better identified by comparing the current post in relation to previous *and* future posts made by u, and by the recent work of [36] who explored bi-directional neural ordinary differential equations for classifying posts on social media, we further extend the HEAT representations to be bi-directional: we learn a HEAT representation over historical posts with timestamps $t < t^{(i)}$, which is then concatenated with another HEAT representation that is learned over future posts with timestamps $t > t^{(i)}$.

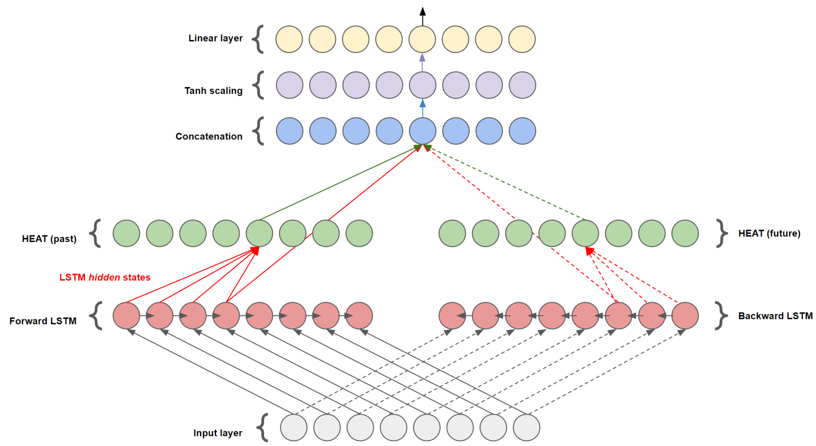

Fig. 1. Architecture of our proposed BiLSTM-HEAT model. Forward and backward LSTM hidden states are aggregated using HEAT (equation 1), in the past and future directions respectively. Future and past HEAT representations are then concatenated for each post index, along with the associated original BiLSTM hidden representation at each post index. After scaling the concatenated representation, this is passed into the final linear layer to predict the label of each post in a given timeline.

Final model. We concatenate the HEAT representations with the original hidden representations of the BiLSTM learned at i. This is to preserve the sequence information captured by the BiLSTM, and linguistic information in the original post embeddings, and to contrast these to the temporal information in the historical and future post representations that are concatenated. We scale all 3 concatenated representations by passing them through a tanh activation function before feeding them as input to the final linear layer, making one prediction per time-step (post) for an entire timeline (see Fig. 1).

3 Experimental Setting

Table 1. Summary of the datasets used in our work.

	Reddit	TalkLife
Users	186	500
Timelines (posts)	255 (6,195)	500 (18,702)
Label distr. % (O/E/S)	77.6 / 15.8 / 6.6	84.5 / 10.8 / 4.7
Timeline Length	$\sim$ 2 months	$\leq$ 2 weeks

Datasets. We work with two datasets containing timelines of longitudinal user posts in English annotated for MoCs (see Table 1). We note that these are the only two such datasets available for this task. These datasets include (a) Reddit, from the CLPsych 2022 shared task [37]; and (b) TalkLife [38], a social network for mental health support. Both datasets were annotated using annotation guidelines introduced in [38], which are also summarized in our appendix A. (a) was annotated by 4 English (2 native) speakers and (b) by 3 English speaking (1 native) university educated annotators. We keep the train/test split used by [37] for (a); for (b) we perform the same 5-fold cross-validation as in [38].

Models, Baselines and Metrics. We represent each post in a timeline as its `[CLS]` representation extracted from BERT [19] fine-tuned for identifying MoCs at the post-level on our training data, using focal loss [26]. On **Reddit**, we contrast our performance against state-of-the-art (SOTA) from the CLPsych 2022 shared task: (a) UoS [4] is an attention-based BiLSTM operating on different input representations of each post of a timeline; (b) WResearch [7] is a XGBoost classifier, fed with emotionally-informed and abnormality seq2seq-based vectors for each post. On **TalkLife**, we compare against SOTA from [38]: (a) BERT(f) is a post-level BERT-based classifier, trained using focal loss; (b) BiLSTM-bert is a timeline-level BiLSTM operating on the posts, which are represented as the `[CLS]` token of (a). We report (per-class/macro-averaged) precision, recall and F1 scores (i) on the test set of Reddit, and (ii) on the 5 folds of TalkLife (macro-averaged). The grid-searched hyper-parameters and training details are provided in Appendix B.

4 Results

4.1 Comparison against SOTA

Results of our model (BiLSTM-HEAT) and baselines in Reddit and TalkLife are shown in Table 3. BiLSTM-HEAT surpasses all models on Reddit in nearly all evaluation metrics and classes, offering a 5.7% relative improvement on macro-F1 compared to current SOTA – UoS [4]. Importantly, it achieves a large performance gain of 17.6% relative on F1 over UoS – also a BiLSTM-based model – on the most challenging class (switch), highlighting the importance of time-aware modelling for capturing rare cases (see Table 1).

On TalkLife, BiLSTM-HEAT fails to outperform the SOTA BiLSTM-bert in most cases, which could be due to the difference in temporal granularity between the TalkLife and Reddit datasets – where Reddit consists of much longer timelines with much larger time-intervals between posts. As a result, modelling the heterogeneity in the time-differences between posts may be more important for identifying MoCs for this dataset – whereas modelling sequence dynamics alone is sufficient for timelines with shorter time-intervals between posts, which we investigate further in our ablation study, presented next.

Table 2. Per-class and macro-averaged results on each dataset (Reddit, TalkLife). **Best** scores are highlighted.

Reddit	**macro-avg**			**S**			**E**			**O**		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Majority	–	.333	.280	–	.000	.000	–	.000	.000	.724	**1.000**	.840
WResearch	.625	.579	.598	.362	.256	.300	.646	.553	.596	.868	.929	.897
UoS	.689	.625	.649	**.490**	.305	.376	.697	.630	.662	.881	.940	.909
BiLSTM-HEAT	**.706**	**.670**	**.686**	.475	**.415**	**.442**	**.741**	**.654**	**.694**	**.902**	.942	**.921**

TalkLife	**macro-avg**			**S**			**E**			**O**		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Majority	–	.333	.280	–	.000	.000	–	.000	.000	.845	**1.000**	.916
BERT(f)	.520	.554	.534	.260	**.321**	.287	.401	**.478**	.436	.898	.864	.881
BiLSTM-bert	**.621**	**.553**	**.580**	**.397**	.264	**.316**	**.568**	.461	**.508**	**.898**	.936	**.917**
BiLSTM-HEAT	.584	.552	.566	.329	.290	.308	.524	.448	.483	.897	.920	.908

4.2 Ablation Analysis

To study the contribution of each component in our model, we ablate each of the components presented in §2 and train and evaluate the resulting models on both datasets: (a) we keep HEAT but remove the bidirectionality component, so that HEAT operates on the previous hidden states of an LSTM (**-BDR**); (b) we remove the 'modelling sequence dynamics' component (**-MSD**), so that HEAT operates on our raw input embeddings instead of the hidden states – concatenating past and future HEAT representations of $v^{(u,a,i)}$ only.

Table 3 summarises our results. Removing the LSTM component (-MSD) has the worst performance by a large margin, highlighting the importance of sequential modelling for our task. This further illustrates the benefit of applying HEAT over LSTM hidden states, rather than on raw posts – as we are able to take advantage of the dynamics learned by these sequence based models.

Bi-directionality (i.e. applying HEAT both on historical and future directions, compared to only applying HEAT in the historical direction) seems to benefit our model most for predicting escalations, and less so for switches. We attribute this due to escalations being more gradual and smoother changes over several future posts, as opposed to switches which are more abrupt and can be more immediately seen just by considering a shift from the previous context – and as such may benefit less from considering the context in a user's future posts.

Sequence length. We analyse the performance of our model and its variants introduced in our ablation study, investigating the ability of our model to account for the heterogeneity in time-intervals between posts, as well as being able to model sequences of varying lengths.

Figure 2 shows our full model BiLSTM-HEAT is able to best capture sequences of varying lengths, being best suited for capturing moments of change in

Table 3. Ablation study: Scores when removing/altering parts of our model. Results are averaged over 3 different random seeds. **Best** scores are highlighted.

Reddit	**macro-avg**			**S**			**E**			**O**		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BiLSTM-HEAT	.706	.670	.686	.475	.415	.442	**.741**	**.654**	**.694**	**.902**	.942	.921
-BDR	**.722**	**.673**	**.693**	**.533**	**.439**	**.480**	.735	.630	.677	.899	**.950**	**.924**
-MSD	.571	.565	.566	.182	.228	.201	.655	.591	.621	.877	.875	.876

TalkLife	**macro-avg**			**S**			**E**			**O**		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
BiLSTM-HEAT	.584	**.552**	**.566**	**.329**	**.290**	**.308**	.524	**.448**	**.483**	**.897**	.920	.908
-BDR	**.585**	.528	.551	.325	.257	.286	**.540**	.397	.457	.890	**.930**	**.910**
-MSD	.480	.442	.455	.200	.139	.161	.368	.272	.312	.872	.915	.893

timelines with both few and many longitudinal posts. Removing the component which models sequence dynamics (-MSD) leads to the largest performance drop across all sequence lengths, suggesting that modelling LSTM hidden states with exponential time-decay provided by HEAT provides the largest performance gain compared to modelling the raw posts without considering the sequential dynamics. Indeed for TalkLife, we see a large 23.9% and 10.1% relative performance gain when averaging the macro-average F1 scores from using BiLSTM-HEAT over -MSD for both long ($80 \leq$ posts ≤ 100) and short ($0 \leq$ posts ≤ 20) sequences. Modelling posts in a bi-directional manner also leads to another performance gain of 1.5% and 6.6% when comparing BiLSTM-HEAT to -BDR by averaging the macro-average F1 scores for the same long and short sequences. This further suggests that considering the user's representations of their future and past together is well suited for identifying changes in a user.

For Reddit a much higher relative performance gain over -BDR occurs for shorter sequence lengths, and a lower relative gain for longer sequences is observed – whereas for TalkLife the opposite is true. Furthermore, we also see that performance decreases with sequence length for all models on Reddit, but conversely performance increases for all models (except -MSD) on TalkLife. We attribute these due to differences in the domains of the datasets, where posts on Reddit are specifically made on mental health related subreddits which might be more indicative of a MoC, whereas posts on TalkLife are more general – discussing day-to-day-life. This is also supported by the easier nature of the task on Reddit based on our results in table 3.

Time-intervals. Figure 3 demonstrates BiLSTM-HEAT outperforms all metrics for varying time intervals relative to previous posts, for nearly all classes and metrics on TalkLife – slightly under performing -BDR only in macro-average precision and precision for escalations. On Reddit, BiLSTM-HEAT achieves the best macro average F1 score for capturing the rare class "escalation" overall across varying time-intervals, with a relative performance gain in F1 of 1.4% and 9.0%

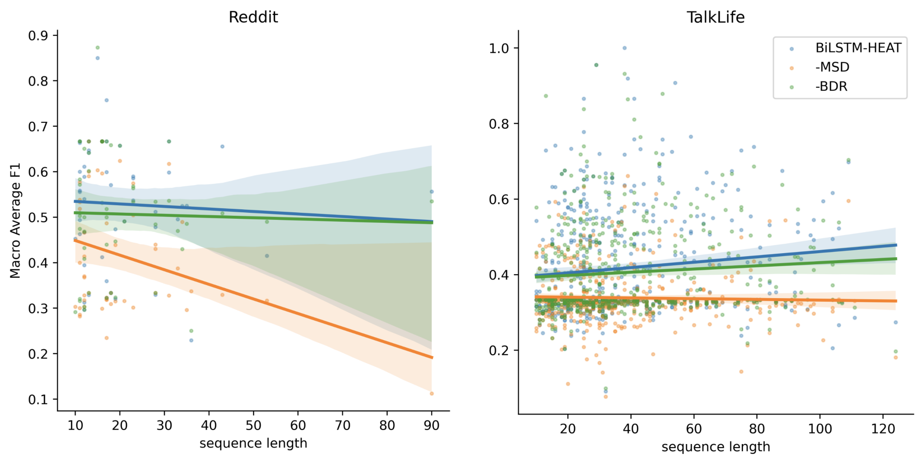

Fig. 2. Performance of models when assessing posts in timelines of varying sequence lengths.

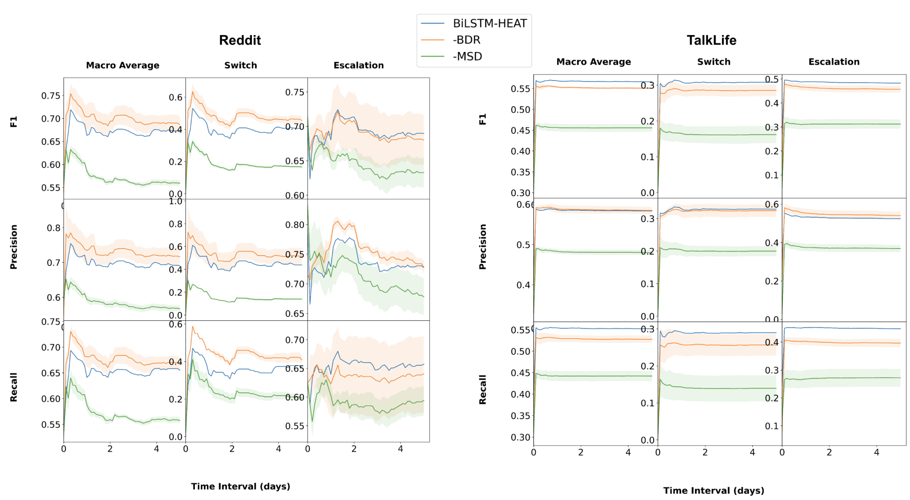

Fig. 3. Effect of time-intervals when assessing posts for MoCs, with the models presented in our ablation study. We evaluate the performance on posts that have a time-interval of less than or equal to what is presented on the x-axis.

on larger timer intervals ($\leq$ 5 days). On Reddit our model struggles to model switches effectively – being outperformed by -BDR, which only considers the forward direction when modelling sequence dynamics and performing aggregation with HEAT. -BDR observed a gain in F1 over BiLSTM of 28.9% on smaller time intervals and a lower relative gain of 14.1% on larger time intervals. This further suggests that for Reddit, the more abrupt class (switch) is better captured by only considering the immediate previous context rather than the rest of the user's distant future posts with bi-directionality, as discussed earlier from analysing table 3. Compared to -MSD, we see a high 81.4% relative performance gain in F1-score for BiLSTM-HEAT on Reddit for smaller time-intervals ($\leq$ 1 day) and a significantly higher 240.0% relative gain on longer time-intervals ($\leq$ 5 days), for assessing "switches" – demonstrating the effectiveness of modelling BiLSTM hidden states with HEAT in a bi-directional manner.

5 Conclusion

In this paper we have presented a time-sensitive approach for building representations of users at different points in time, based on linguistic and temporal context in social media posts in a bi-directional manner. By modelling a user's timeline of posts with a BiLSTM parametrized by time-dependent exponential decay with a Hawkes process, we demonstrate superior performance over prior approaches which did not consider temporality and bi-directionality together when assessing changes in mood of individuals – outperforming the best performing systems on the CLPsych 2022 shared task, which targets the same objective.

6 Limitations

Our model is trained to predict the presence of MoCs on the basis of content shared online by social media users. As such, it cannot generalise to detect changes in mood unless these are reflected in an individuals' posts. This fact has a further downstream effect in sample bias, since our datasets consist of users who have (a) certain demographic characteristics as social media users who post in English and (b) have *selected* to self-disclose their well-being online, which has been shown to lack generalisability in related (user-level) mental-health tasks [21].

The limitation of generalisability is also present in the ability of our models to effectively predict MoCs in datasets of different characteristics. For example, the BERT(f) model trained on TalkLife, which achieved the second-best results in [38] on this dataset, was easily outperformed when applied on Reddit by a simple post-level logistic regression trained with in-domain tfidf feature vectors [37]. This highlights the importance of the different characteristics of each platform (e.g., length/content of messages) as well as of the different time intervals used to define what constitutes a 'user timeline' in the pre-annotation stage (e.g. 2 weeks in TalkLife *vs* 2 months on Reddit).

Finally, the definitions of MoC that we have followed in our work has been established on the basis of *mood changes in social media*. A well-established model performance on other related types of NLP-based MoC identification tasks, such as detecting changes during psychotherapy sessions, cannot be safely hypothesised from the findings of this work.

7 Ethics Statement

Ethics IRB approval was obtained from the IRB Committee of the lead University prior to engaging in this research study. Our work involves ethical considerations around the analysis of user generated content shared on social media (TalkLife and Reddit). A license was obtained to work with the user data from TalkLife and a project proposal was submitted to them in order to embark on the project. Potential risks from the application of our work in being able to identify moments of change in individuals' timelines are akin to the identification of those in earlier work on personal event identification from social media and the detection of suicidal ideation. Potential mitigation strategies include restricting and regulating access to the code base and annotation labels used for evaluation.

References

1. Alhamed, F., Ive, J., Specia, L.: Predicting moments of mood changes overtime from imbalanced social media data. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 239–244. Association for Computational Linguistics, Seattle, USA (Jul 2022). https://doi.org/10.18653/v1/2022.clpsych-1.23, https://aclanthology.org/2022.clpsych-1.23
2. AlHamed, F., Ive, J., Specia, L.: Predicting moments of mood changes overtime from imbalanced social media data. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 239–244 (2022)
3. Amir, S., Coppersmith, G., Carvalho, P., Silva, M.J., Wallace, B.C.: Quantifying mental health from social media with neural user embeddings. In: Machine Learning for Healthcare Conference. pp. 306–321. PMLR (2017)
4. Azim, T., Gyanendro Singh, L., Middleton, S.E.: Detecting moments of change and suicidal risks in longitudinal user texts using multi-task learning. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 213–218. Association for Computational Linguistics, Seattle, USA (Jul 2022). https://doi.org/10.18653/v1/2022.clpsych-1.19, https://aclanthology.org/2022.clpsych-1.19
5. Azim, T., Singh, L., Middleton, S.: Detecting moments of change and suicidal risks in longitudinal user texts using multi-task learning. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 213–218 (2022)
6. Bagroy, S., Kumaraguru, P., De Choudhury, M.: A social media based index of mental well-being in college campuses. In: Proceedings of the 2017 CHI Conference on Human factors in Computing Systems. pp. 1634–1646 (2017)
7. Bayram, U., Benhiba, L.: Emotionally-informed models for detecting moments of change and suicide risk levels in longitudinal social media data. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 219–225. Association for Computational Linguistics, Seattle, USA (Jul 2022). https://doi.org/10.18653/v1/2022.clpsych-1.20, https://aclanthology.org/2022.clpsych-1.20
8. Bayram, U., Benhiba, L.: Emotionally-informed models for detecting moments of change and suicide risk levels in longitudinal social media data. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 219–225 (2022)
9. Boinepelli, S., Subramanian, S., Singam, A., Raha, T., Varma, V.: Towards capturing changes in mood and identifying suicidality risk. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 245–250. Association for Computational Linguistics, Seattle, USA (Jul 2022). https://doi.org/10.18653/v1/2022.clpsych-1.24, https://aclanthology.org/2022.clpsych-1.24
10. Boinepelli, S., Subramanian, S., Singam, A., Raha, T., Varma, V.: Towards capturing changes in mood and identifying suicidality risk. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 245–250 (2022)
11. Bucur, A.M., Jang, H., Liza, F.F.: Capturing changes in mood over time in longitudinal data using ensemble methodologies (2022)

12. Coppersmith, G., Dredze, M., Harman, C., Hollingshead, K., Mitchell, M.: CLPsych 2015 shared task: Depression and PTSD on Twitter. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. pp. 31–39. Association for Computational Linguistics, Denver, Colorado (Jun 5 2015). https://doi.org/10.3115/v1/W15-1204, https://aclanthology.org/W15-1204
13. Coppersmith, G., Ngo, K., Leary, R., Wood, A.: Exploratory analysis of social media prior to a suicide attempt. In: Proceedings of the third workshop on computational linguistics and clinical psychology. pp. 106–117 (2016)
14. Culnan, J., Diaz, D.R., Bethard, S.: Exploring transformers and time lag features for predicting changes in mood over time. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 226–231 (2022)
15. Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure. Springer (2008)
16. Daley, D.J., Vere-Jones, D., et al.: An introduction to the theory of point processes: volume I: elementary theory and methods. Springer (2003)
17. Daly, M., Sutin, A.R., Robinson, E.: Longitudinal changes in mental health and the COVID-19 pandemic: evidence from the UK Household Longitudinal Study. Psychological Medicine pp. 1–10 (Nov 2020). https://doi.org/10.1017/S0033291720004432, https://www.cambridge.org/core/product/identifier/S0033291720004432/type/journal_article
18. De Choudhury, M., Gamon, M., Counts, S., Horvitz, E.: Predicting depression via social media. In: Proceedings of the International AAAI Conference on Web and Social Media. vol. 7 (2013)
19. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. pp. 4171–4186 (Jun 2019). https://doi.org/10.18653/v1/N19-1423, https://www.aclweb.org/anthology/N19-1423
20. Du, N., Dai, H., Trivedi, R., Upadhyay, U., Gomez-Rodriguez, M., Song, L.: Recurrent marked temporal point processes: Embedding event history to vector. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp. 1555–1564 (2016)
21. Ernala, S.K., Birnbaum, M.L., Candan, K.A., Rizvi, A.F., Sterling, W.A., Kane, J.M., De Choudhury, M.: Methodological gaps in predicting mental health states from social media: triangulating diagnostic signals. In: Proceedings of the 2019 chi conference on human factors in computing systems. pp. 1–16 (2019)
22. Gamaarachchige, P.K., Orabi, A.H., Orabi, M.H., Inkpen, D.: Multi-task learning to capture changes in mood over time. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 232–238 (2022)
23. Ganesan, A.V., Varadarajan, V., Mittal, J., Subrahmanya, S., Matero, M., Soni, N., Guntuku, S.C., Eichstaedt, J., Schwartz, H.A.: Wwbp-sqt-lite: Multi-level models and difference embeddings for moments of change identification in mental health forums. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 251–258 (2022)
24. Gkotsis, G., Oellrich, A., Velupillai, S., Liakata, M., Hubbard, T.J., Dobson, R.J., Dutta, R.: Characterisation of mental health conditions in social media using informed deep learning. Scientific reports **7**(1), 1–11 (2017)
25. Hawkes, A.G.: Point spectra of some mutually exciting point processes. Journal of the Royal Statistical Society: Series B (Methodological) **33**(3), 438–443 (1971)

26. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal Loss for Dense Object Detection. pp. 2980–2988 (2017), https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html
27. Loveys, K., Crutchley, P., Wyatt, E., Coppersmith, G.: Small but mighty: affective micropatterns for quantifying mental health from social media language. In: Proceedings of the fourth workshop on computational linguistics and clinical Psychology—From linguistic signal to clinical reality. pp. 85–95 (2017)
28. Marcos, H.F., Cejudo, A., Martinez-Romo, J., Pérez, A., Araujo, L., Lebea, N., Oronoz, M., Casillas, A.: Approximate nearest neighbour extraction techniques and neural networks for suicide risk prediction in the clpsych 2022 shared task. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 199–204 (2022)
29. Masuda, N., Kurahashi, I., Onari, H.: Suicide ideation of individuals in online social networks. PloS one **8**(4), e62262 (2013)
30. Preoţiuc-Pietro, D., Eichstaedt, J., Park, G., Sap, M., Smith, L., Tobolsky, V., Schwartz, H.A., Ungar, L.: The role of personality, age, and gender in tweeting about mental illness. In: Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality. pp. 21–30 (2015)
31. Rizoiu, M.A., Lee, Y., Mishra, S., Xie, L.: A tutorial on hawkes processes for events in social media. arXiv preprint arXiv:1708.06401 (2017)
32. Sawhney, R., Agarwal, S., Wadhwa, A., Shah, R.: Exploring the Scale-Free Nature of Stock Markets: Hyperbolic Graph Learning for Algorithmic Trading. In: Proceedings of the Web Conference 2021. pp. 11–22. ACM, Ljubljana Slovenia (Apr 2021). https://doi.org/10.1145/3442381.3450095, https://dl.acm.org/doi/10.1145/3442381.3450095
33. Sawhney, R., Joshi, H., Shah, R.R., Flek, L.: Suicide Ideation Detection via Social and Temporal User Representations using Hyperbolic Learning. In: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. pp. 2176–2190. Association for Computational Linguistics, Online (2021). https://doi.org/10.18653/v1/2021.naacl-main.176, https://www.aclweb.org/anthology/2021.naacl-main.176
34. Shchur, O., Türkmen, A.C., Januschowski, T., Günnemann, S.: Neural temporal point processes: A review. arXiv preprint arXiv:2104.03528 (2021)
35. Shing, H.C., Nair, S., Zirikly, A., Friedenberg, M., Daumé III, H., Resnik, P.: Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. pp. 25–36 (2018)
36. Tamire, M., Anumasa, S., Srijith, P.K.: Bi-directional recurrent neural ordinary differential equations for social media text classification. In: Proceedings of the 2nd Workshop on Deriving Insights from User-Generated Text. p. 20–24. Association for Computational Linguistics, (Hybrid) Dublin, Ireland, and Virtual (May 2022). https://doi.org/10.18653/v1/2022.wit-1.3, https://aclanthology.org/2022.wit-1.3
37. Tsakalidis, A., Chim, J., Bilal, I.M., Zirikly, A., Atzil-Slonim, D., Nanni, F., Resnik, P., Gaur, M., Roy, K., Inkster, B., Leintz, J., Liakata, M.: Overview of the CLPsych 2022 shared task: Capturing moments of change in longitudinal user posts. In: Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology. pp. 184–198. Association for Computational Lin-

guistics, Seattle, USA (Jul 2022). https://doi.org/10.18653/v1/2022.clpsych-1.16, https://aclanthology.org/2022.clpsych-1.16
38. Tsakalidis, A., Nanni, F., Hills, A., Chim, J., Song, J., Liakata, M.: Identifying moments of change from longitudinal user text. In: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 4647–4660. Association for Computational Linguistics, Dublin, Ireland (May 2022). https://doi.org/10.18653/v1/2022.acl-long.318, https://aclanthology.org/2022.acl-long.318

A Annotation Guidelines

The two datasets of longitudinal user posts annotated for MoCs that we make use of in this paper were sourced by [38] and [37] for TalkLife and Reddit respectively. Both datasets were annotated using the same annotation guidelines and annotation interface proposed [38].

Annotators were provided with timelines to view, containing chronologically ordered posts by users, along with their associated comments and timestamps. They were then asked to label posts for MoCs.

The first type of label, "Switch" was defined in the guidelines as a "drastic change in mood, in comparison with the recent past". Annotators were also tasked to label how long the Switch in mood persists (i.e. label its beginning and end). The second type of label "Escalation" was defined in their guidelines as a "gradual change in mood, which should last for a few posts". Similarly, annotators were also instructed to label the associated range of posts for how long this change persists: where a peak of the escalation must be labelled, and the beginning and end of the gradual mood change also provided. Finally, a label of "None" was provided by default where no mood change was identified for that given post.

B Hyper-parameters Searched

We perform a grid-search over the HEAT parameters (eq 1): both β (decay rate) and α (self-excitation) in the range $[0.00001, 0.001, 0.1]$ for both datasets. All models are searched with learning rates in the range $[0.0001, 0.001, 0.01]$ on Reddit and $[0.001, 0.01]$ for TalkLife. All models are trained with 100 epochs with early stopping using a patience of 5 for all models and both datasets. For the BiLSTM module, we perform a grid-search over all layers using output dimensions of $[128, 256, 512]$ and $[128, 256]$ for Reddit and Talklife respectively. All models were implemented with PyTorch, and were trained using K-Fold cross validation over 5 folds using training, validation, and testing sizes of 60%, 20%, 20% respectively.