# Dynamic Representations for Longitudinal and Temporal Language Modelling

Talia Tseriotou

The Alan Turing Institute
February 19th 2024

# Research Motivation

- LLMs are powerful in producing **static** word embeddings.
- **Limited work** on dynamic user representations.
- **Current research**:
  - applied only off-line.
  - lacks generalisability.

# Research Motivation

- LLMs are powerful in producing **static** word embeddings.
- **Limited work** on dynamic user representations.
- **Current research**:
  - applied only off-line.
  - lacks generalisability.

**Build efficient and compressed temporal user representations to address user-specific changes over time**

# Task Overview

**Identify Moments of Change in Longitudinal User Social Media Data**

# Task Overview

**Timeline** posts annotation to identify moments of change (MoC):

- **Switches (IS)**: sudden mood shifts from 😊 to 😔 or vice versa
- **Escalations (IE)**: gradual mood progression from

  😒 or 😊 → 🤩 or

  😒 or 😔 → 😢
- **None (O)**: No change

# Timeline Example

*“Gonna be offline for a bit, having lunch atm”*

**Timeline**

**O**

$t_n$

# Timeline Example

*"Gonna be offline for a bit, having lunch atm"*

*"OK, I am back. Kinda bored."*

**Timeline**

| O | O |
|---|---|
| $t_n$ | $t_{n+1}$ |

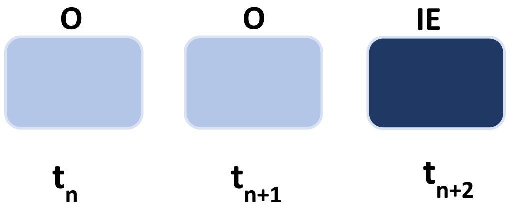
# Timeline Example

*"Gonna be offline for a bit, having lunch atm"*

*"OK, I am back. Kinda bored."*

*"Everything is just so wrong in my life"*

## Timeline



# Timeline Example

*“Gonna be offline for a bit, having lunch atm”*

*“OK, I am back. Kinda bored.”*

*“Everything is just so wrong in my life”*

*“Omg can’t stop crying, everything is ruined”*

**Timeline**

| O | O | IE | IE |
|---|---|---|---|
| $t_n$ | $t_{n+1}$ | $t_{n+2}$ | $t_{n+3}$ |

# Timeline Example

*“Gonna be offline for a bit, having lunch atm”*

*“OK, I am back. Kinda bored.”*

*“Everything is just so wrong in my life”*

*“Omg can’t stop crying, everything is ruined”*

*“Need someone before I do something stupid !PLEASE HELP!”*

## Timeline

| O | O | IE | IE | IE Peak |
|---|---|---|---|---|
| $t_n$ | $t_{n+1}$ | $t_{n+2}$ | $t_{n+3}$ | $t_{n+4}$ |

# Timeline Example

"Gonna be offline for a bit, having lunch atm"

"OK, I am back. Kinda bored."

"Everything is just so wrong in my life"

"Omg can't stop crying, everything is ruined"

"Need someone before I do something stupid !PLEASE HELP!"

"Wish things were differently.. Miss my gf.."

## Timeline

| O | O | IE | IE | IE Peak | IE |
|---|---|---|---|---|---|
| $t_n$ | $t_{n+1}$ | $t_{n+2}$ | $t_{n+3}$ | $t_{n+4}$ | $t_{n+5}$ |

# Timeline Example

"Gonna be offline for a bit, having lunch atm"

"OK, I am back. Kinda bored."

"Everything is just so wrong in my life"

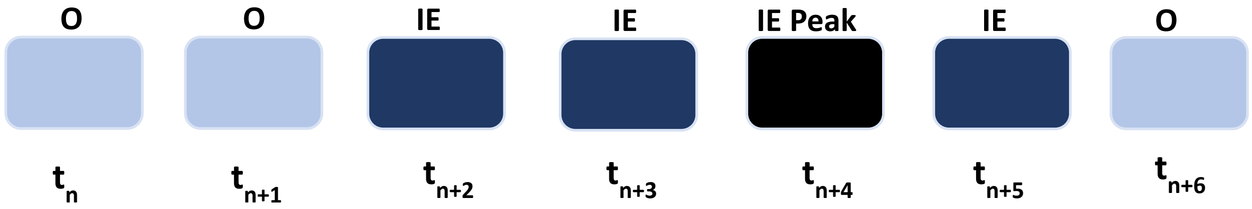
"Omg can't stop crying, everything is ruined"

"Need someone before I do something stupid !PLEASE HELP!"

"Wish things were differently.. Miss my gf.."

"Having an exam in one week, hope to be able to do well"

## Timeline



# Timeline Example

“Gonna be offline for a bit, having lunch atm”

“OK, I am back. Kinda bored.”

“Everything is just so wrong in my life”

“Omg can’t stop crying, everything is ruined”

“Need someone before I do something stupid !PLEASE HELP!”

“Wish things were differently.. Miss my gf..”

“Having an exam in one week, hope to be able to do well”

“Has anyone watched the last episode of Game of Thrones? Is it any good?”

## Timeline

| O | O | IE | IE | IE Peak | IE | O | O |
|---|---|---|---|---|---|---|---|
| $t_n$ | $t_{n+1}$ | $t_{n+2}$ | $t_{n+3}$ | $t_{n+4}$ | $t_{n+5}$ | $t_{n+6}$ | $t_{n+7}$ |

# Timeline Example

*“Gonna be offline for a bit, having lunch atm”*

*“OK, I am back. Kinda bored.”*

*“Everything is just so wrong in my life”*

*“Omg can’t stop crying, everything is ruined”*

*“Need someone before I do something stupid !PLEASE HELP!”*

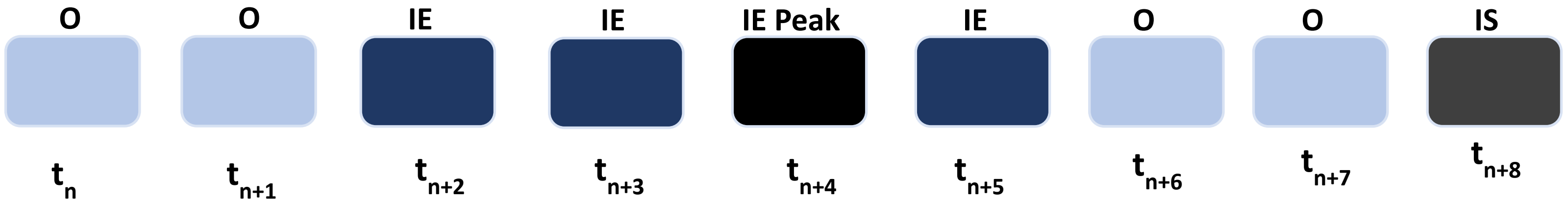
*“Wish things were differently.. Miss my gf..”*

*“Having an exam in one week, hope to be able to do well”*

*“Has anyone watched the last episode of Game of Thrones? Is it any good?”*

*“My partner called, told me she misses me and wants to meet! Best day E-V-E-R!!”*

## Timeline



# Timeline Example

*“Gonna be offline for a bit, having lunch atm”*

*“OK, I am back. Kinda bored.”*

*“Everything is just so wrong in my life”*

*“Omg can’t stop crying, everything is ruined”*

*“Need someone before I do something stupid !PLEASE HELP!”*

*“Wish things were differently.. Miss my gf..”*

*“Having an exam in one week, hope to be able to do well”*

*“Has anyone watched the last episode of Game of Thrones? Is it any good?”*

*“My partner called, told me she misses me and wants to meet! Best day E-V-E-R!!”*

*“Can anyone recommend any decent crime book to read”?*

## Timeline

| O | O | IE | IE | IE Peak | IE | O | O | IS | O |
|---|---|---|---|---|---|---|---|---|---|
| $t_n$ | $t_{n+1}$ | $t_{n+2}$ | $t_{n+3}$ | $t_{n+4}$ | $t_{n+5}$ | $t_{n+6}$ | $t_{n+7}$ | $t_{n+8}$ | $t_{n+9}$ |

# Data

| | TalkLife | Reddit |
|---|---|---|
| Timelines | 500 | 256 |
| Timeline length | ≤ 2-week | ~ 2-month |
| Avg posts per timeline | 37 | 24 |

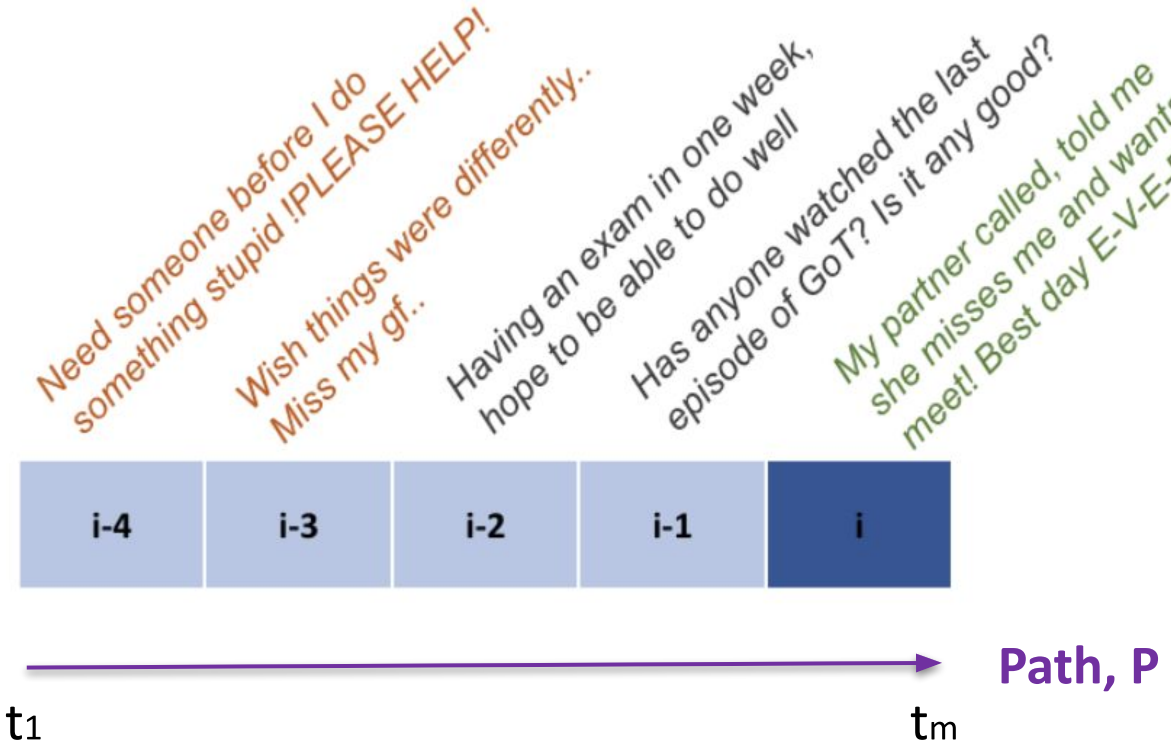
# Path Signatures

# What are Path Signatures?

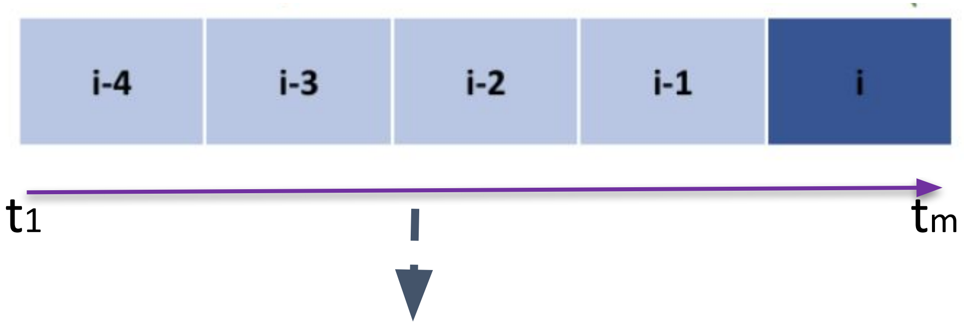
**In theory:** **Path-wise definition to the solution of differential equation driven by irregular signals.**

**In practice:** **Produce a collection of statistics summarizing uniquely important information of the path.**

# Path Signatures - Explained



# Path Signatures - Explained




Path, P

$$S(P)_{t_1,t_m} = (1, S(P)^{1}_{t_1,t_m}, \ldots, S(P)^{c}_{t_1,t_m},$$
$$S(P)^{1,1}_{t_1,t_m}, S(P)^{1,2}_{t_1,t_m}, \ldots, S(P)^{c,c}_{t_1,t_m}$$
$$\ldots, S(P)^{i_1,i_2,\cdots,i_r}_{t_1,t_m}, \ldots)$$

$$S(P)^{i_1,i_2,\cdots,i_r}_{t_1,t_m} = \int_{g_r} \ldots \int_{g_1} dP^{i_1}_{g_1} \otimes \ldots \otimes dP^{i_r}_{g_r}$$

Path Signature, S(P)

# Path Signatures - Explained

$$S(P)_{t_1,t_m} = (1, \boxed{S(P)^1_{t_1,t_m}, \ldots, S(P)^c_{t_1,t_m}},$$

$$S(P)^{1,1}_{t_1,t_m}, S(P)^{1,2}_{t_1,t_m}, \ldots, S(P)^{c,c}_{t_1,t_m}$$

$$\ldots, S(P)^{i_1,i_2,\cdots,i_r}_{t_1,t_m}, \ldots)$$

degree 1

**Path Signature, S(P)**

# Path Signatures - Explained

$$S(P)_{t_1,t_m} = (1, S(P)^1_{t_1,t_m}, ..., S(P)^c_{t_1,t_m},$$

$$\boxed{S(P)^{1,1}_{t_1,t_m}, S(P)^{1,2}_{t_1,t_m}, ..., S(P)^{c,c}_{t_1,t_m}}$$ degree 2

$$..., S(P)^{i_1,i_2,\cdots,i_r}_{t_1,t_m}, ...)$$

**Path Signature, S(P)**

# Path Signatures - Explained

$$S(P)_{t_1,t_m} = (1, S(P)^1_{t_1,t_m}, ..., S(P)^c_{t_1,t_m},$$
$$S(P)^{1,1}_{t_1,t_m}, S(P)^{1,2}_{t_1,t_m}, ..., S(P)^{c,c}_{t_1,t_m}$$
$$..., S(P)^{i_1,i_2,\cdots,i_r}_{t_1,t_m}, ...)$$

Number of output dimensions used as features:

$$(c^{N+1} - c)(c-1)^{-1}$$

**Path Signature, S(P)**

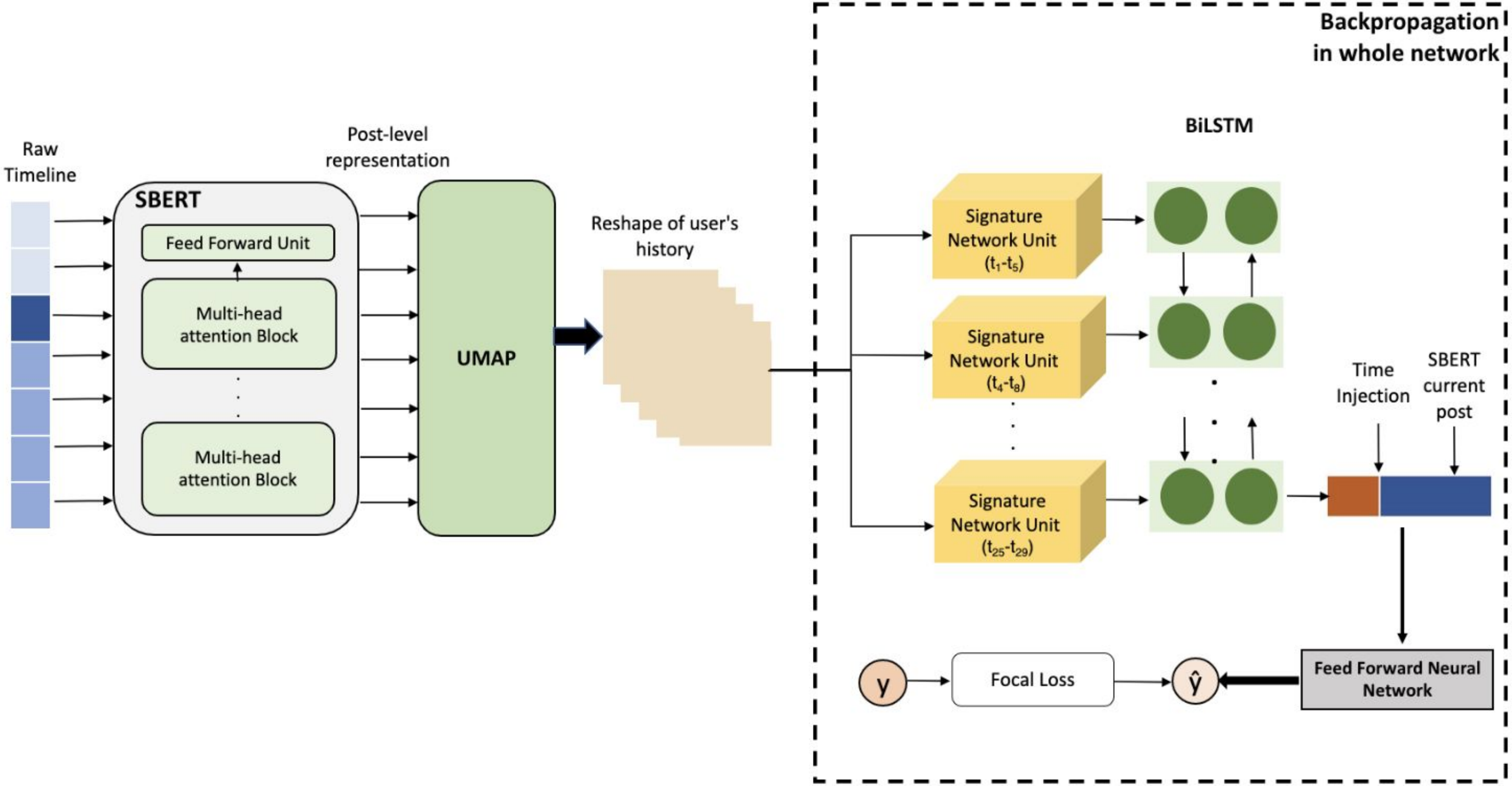
**Dimensionality Reduction**

# Path Signatures

Advantages:

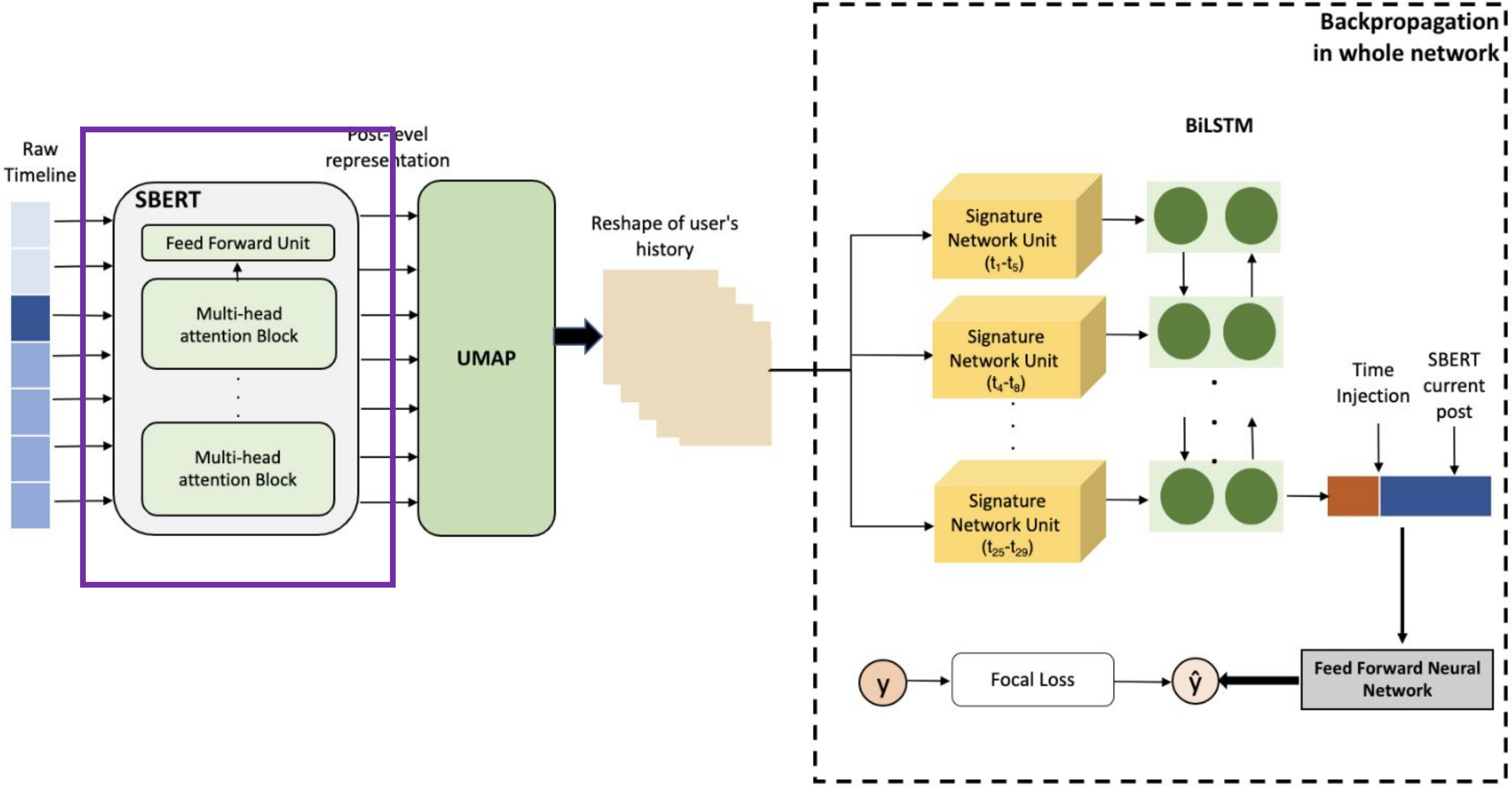
- Efficient and compressed encoding of **sequential data**.
- Sequential **pooling** operator in Neural Models.
- Enhances **short-term dependencies** in linguistic timelines.
- Account for **time irregularities**.

# Model

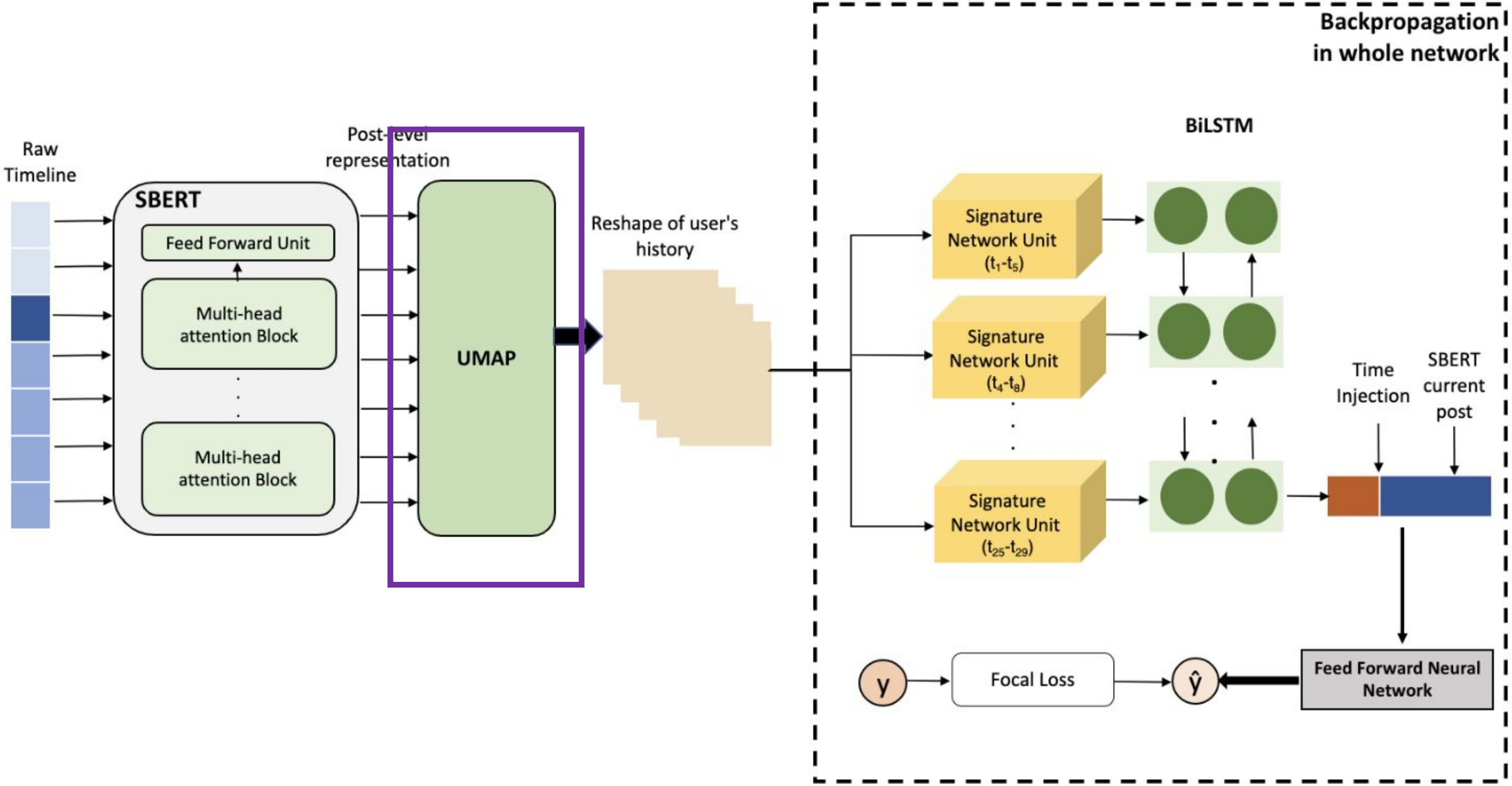


Backpropagation in whole network
Raw Timeline
SBERT
Feed Forward Unit
Multi-head attention Block
Multi-head attention Block
Post-level representation
UMAP
Reshape of user's history
BiLSTM
Signature Network Unit ($t_1$-$t_5$)
Signature Network Unit ($t_4$-$t_8$)
Signature Network Unit ($t_{25}$-$t_{29}$)
Time Injection
SBERT current post
Feed Forward Neural Network
y
Focal Loss
ŷ

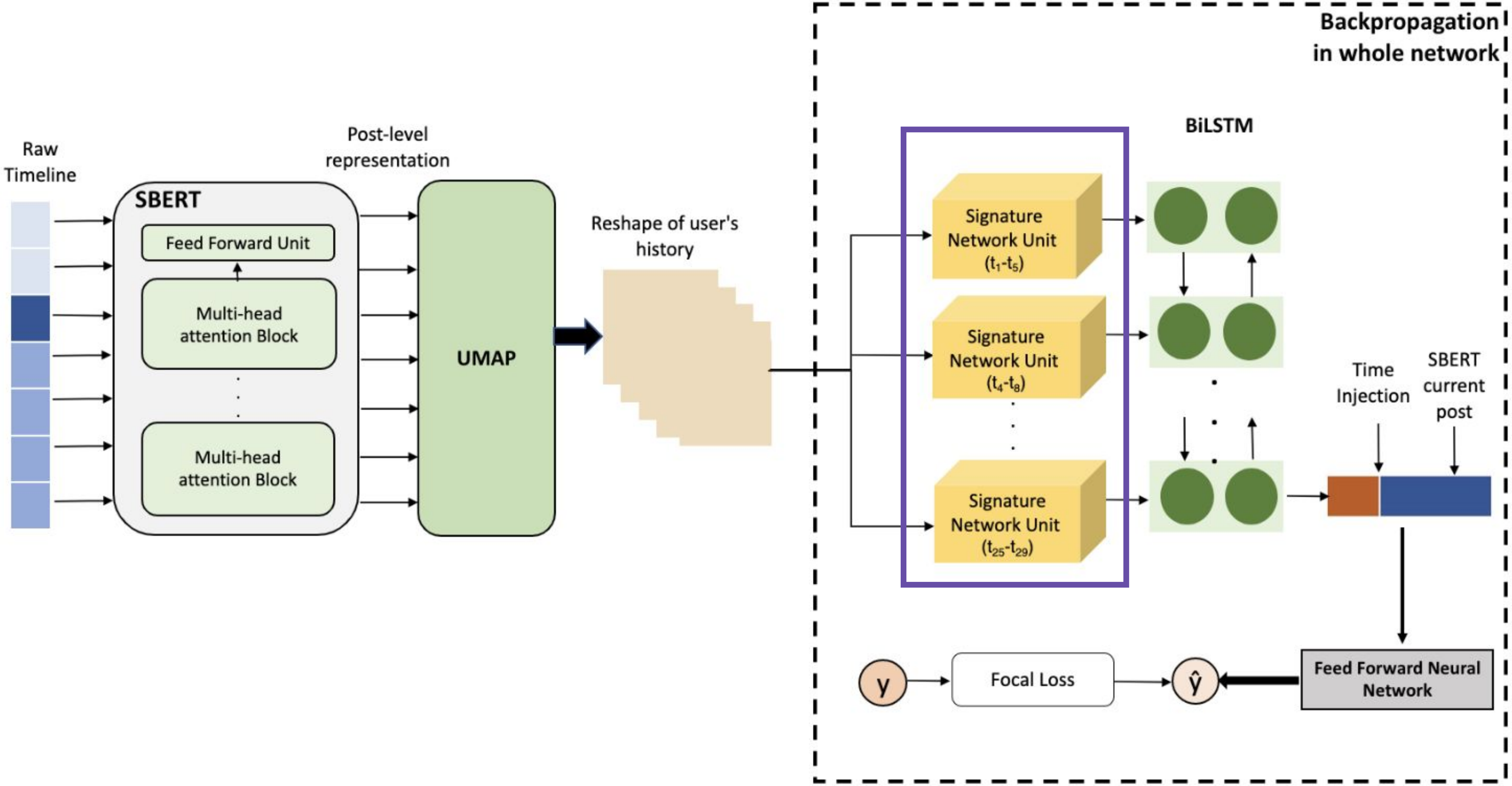
# Model


Backpropagation in whole network
Raw Timeline
Post-level representation
SBERT
Feed Forward Unit
Multi-head attention Block
Multi-head attention Block
UMAP
Reshape of user's history
BiLSTM
Signature Network Unit ($t_1$-$t_5$)
Signature Network Unit ($t_4$-$t_8$)
Signature Network Unit ($t_{25}$-$t_{29}$)
Time Injection
SBERT current post
Feed Forward Neural Network
y
Focal Loss
ŷ

# Model

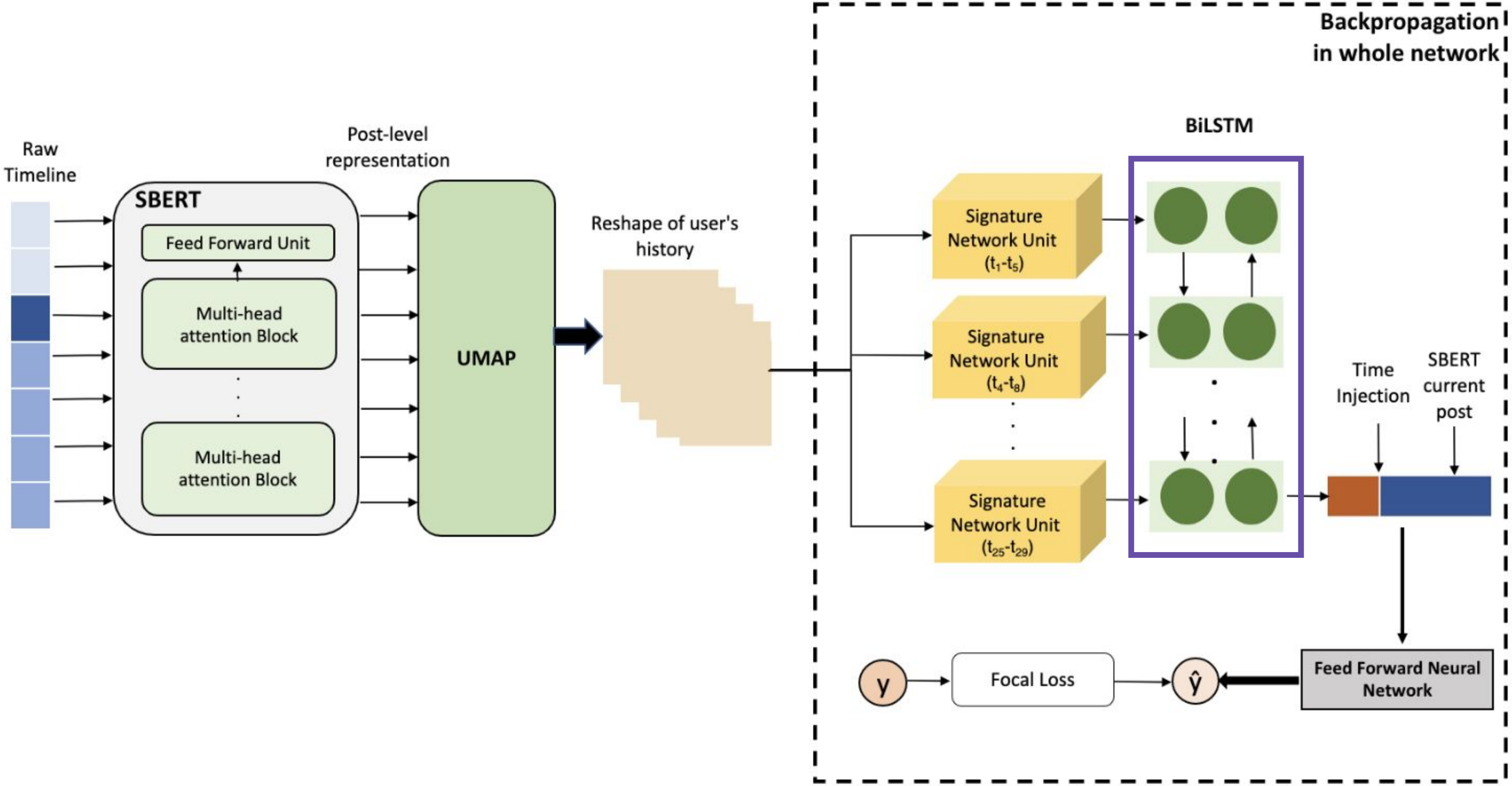

Raw Timeline
SBERT
Feed Forward Unit
Multi-head attention Block
Multi-head attention Block
Post-level representation
UMAP
Reshape of user's history
Backpropagation in whole network
BiLSTM
Signature Network Unit ($t_1$-$t_5$)
Signature Network Unit ($t_4$-$t_8$)
Signature Network Unit ($t_{25}$-$t_{29}$)
Time Injection
SBERT current post
Feed Forward Neural Network
y
Focal Loss
ŷ

# Model



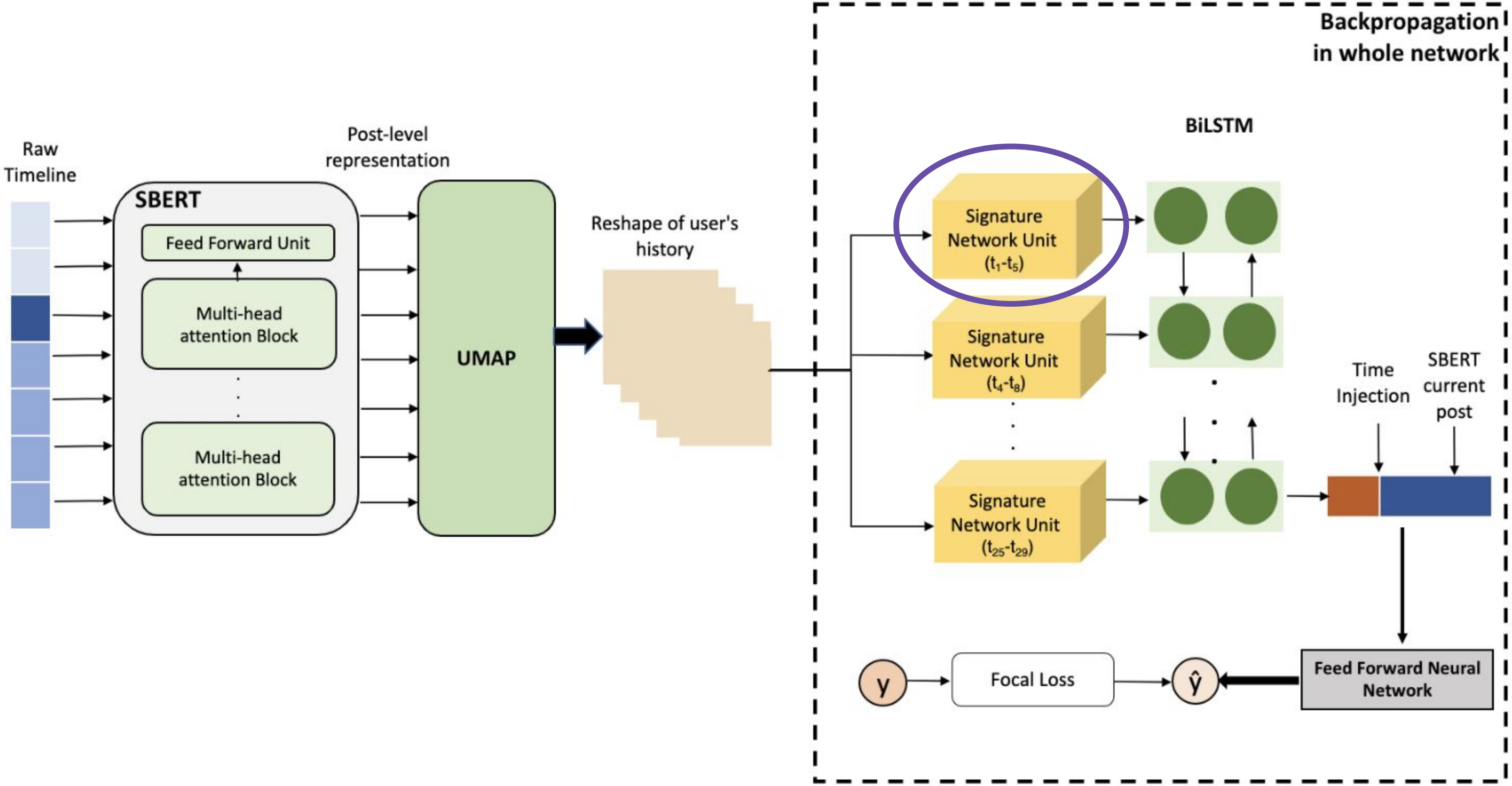
Raw Timeline
SBERT
Feed Forward Unit
Multi-head attention Block
Multi-head attention Block
Post-level representation
UMAP
Reshape of user's history
Backpropagation in whole network
Signature Network Unit ($t_1$-$t_5$)
Signature Network Unit ($t_4$-$t_8$)
Signature Network Unit ($t_{25}$-$t_{29}$)
BiLSTM
Time Injection
SBERT current post
Feed Forward Neural Network
y
Focal Loss
ŷ

# Model



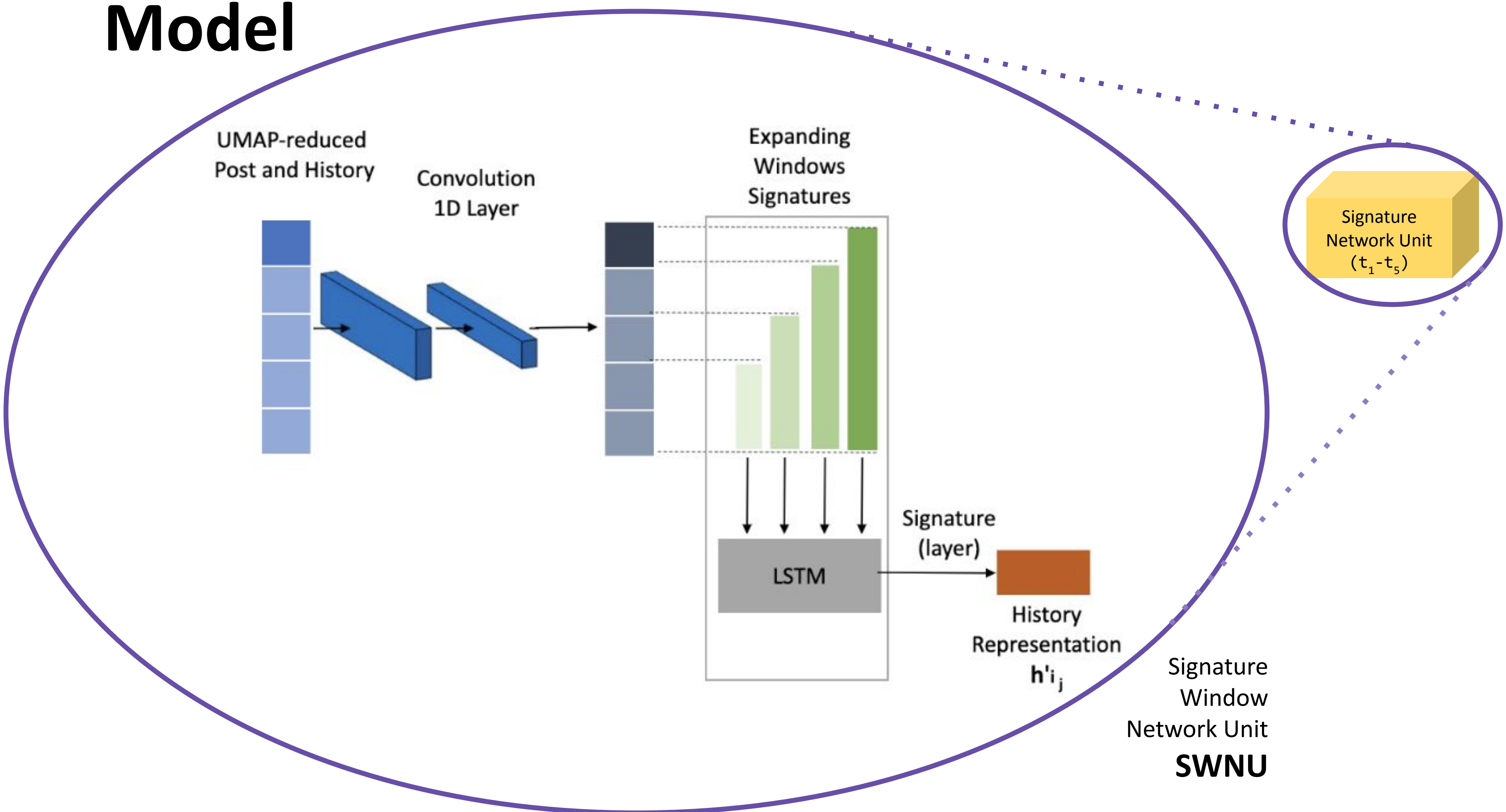
Raw Timeline
SBERT
Feed Forward Unit
Multi-head attention Block
Multi-head attention Block
Post-level representation
UMAP
Reshape of user's history
Backpropagation in whole network
BiLSTM
Signature Network Unit ($t_1$-$t_5$)
Signature Network Unit ($t_4$-$t_8$)
Signature Network Unit ($t_{25}$-$t_{29}$)
Time Injection
SBERT current post
Feed Forward Neural Network
y
Focal Loss
ŷ

# Model



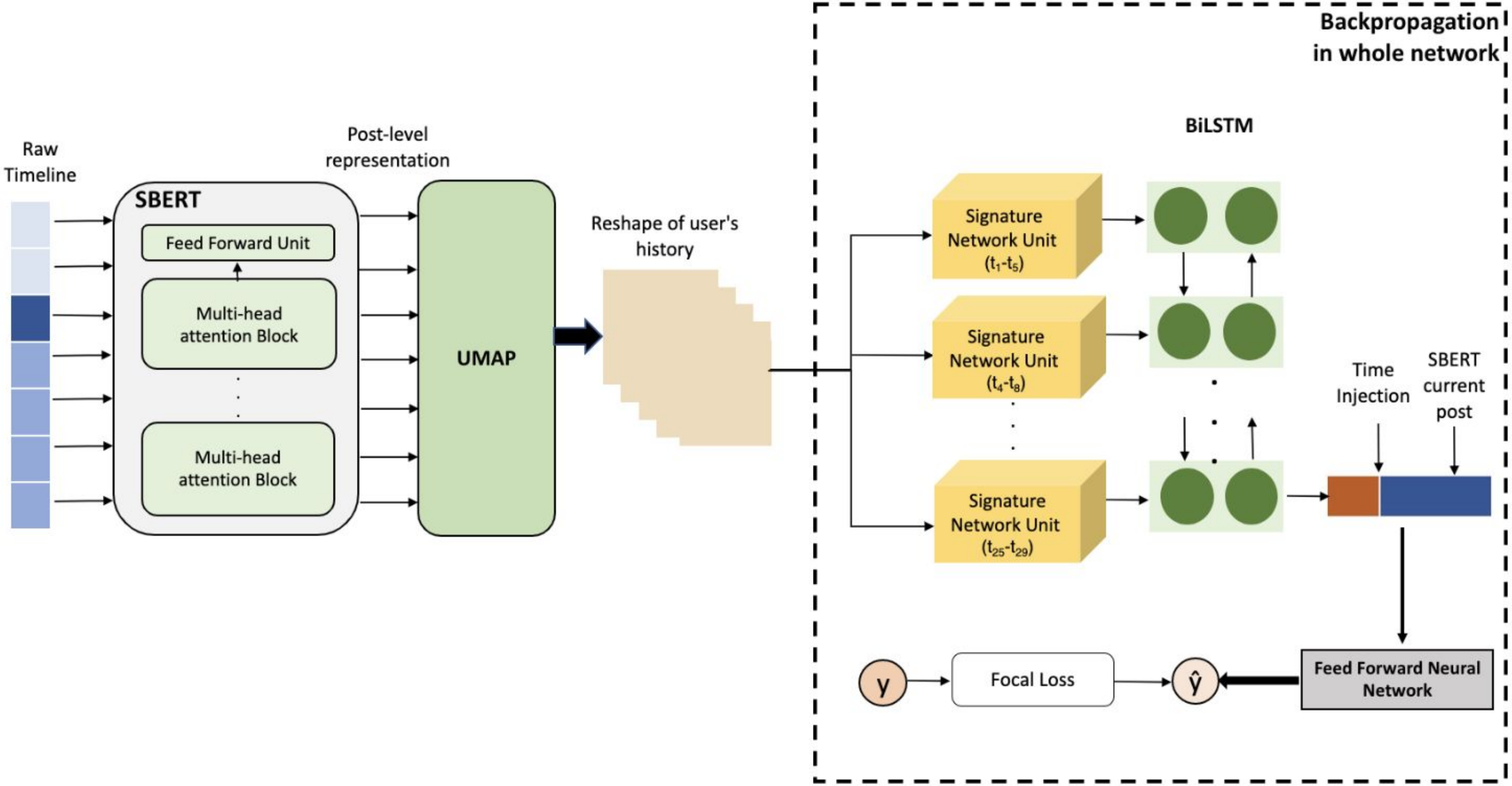
Raw Timeline
SBERT
Feed Forward Unit
Multi-head attention Block
Multi-head attention Block
Post-level representation
UMAP
Reshape of user's history
Backpropagation in whole network
BiLSTM
Signature Network Unit ($t_1$-$t_5$)
Signature Network Unit ($t_4$-$t_8$)
Signature Network Unit ($t_{25}$-$t_{29}$)
Time Injection
SBERT current post
Feed Forward Neural Network
y
Focal Loss
ŷ

# Model




Signature
Window
Network Unit
**SWNU**

# Model



Raw Timeline
SBERT
Feed Forward Unit
Multi-head attention Block
Multi-head attention Block
Post-level representation
UMAP
Reshape of user's history
Backpropagation in whole network
BiLSTM
Signature Network Unit ($t_1$-$t_5$)
Signature Network Unit ($t_4$-$t_8$)
Signature Network Unit ($t_{25}$-$t_{29}$)
Time Injection
SBERT current post
Feed Forward Neural Network
Focal Loss
y
ŷ

# Results

**ACL 2023 (findings)** Sequential Path Signature Networks for Personalised Longitudinal Language Modeling

| | | | IS | | | IE | | | O | | | macro-avg | | | Model Type | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 | Emotion | Future |
| TalkLife | Naïve | Majority | – | – | – | – | – | – | .845 | 1 | .916 | .282 | .333 | .305 | | |
| | | Random | .047 | .047 | .047 | .108 | .108 | .108 | .845 | .845 | .845 | .333 | .333 | .333 | | |
| | Post-level | BERT(f) (Tsakalidis et al., 2022b) | .260 | **.321** | .287 | .401 | .478 | .436 | 898 | .864 | .881 | .520 | .554 | .534 | | |
| | Timeline-level | EM-DM | **.553** | .118 | .193 | .479 | .351 | .405 | .880 | **.948** | .913 | **.631** | .472 | .504 | ✓ | ✓ |
| | (Tsakalidis et al., 2022b) | BiLSTM-bert | .397 | .264 | **.316** | **.568** | .461 | **.508** | .898 | .936 | **.917** | .621 | .553 | **.580** | | ✓ |
| | Timeline-level | SBERT(avg hist) | .283 | .244 | .262 | .424 | .486 | .452 | .896 | .885 | .890 | .534 | .539 | .535 | | |
| | (-signature) | BiLSTM-sbert(hist) | .258 | .272 | .264 | .442 | .506 | .468 | .901 | .879 | .890 | .534 | .553 | .541 | | |
| | | BiLSTM-bert(hist) | .405 | .241 | .302 | .536 | .415 | .468 | .892 | .938 | .914 | .611 | .531 | .561 | | |
| | Timeline-level (+signature) | Seq-Sig-Net (our work) | .331 | .290 | .309 | .435 | **.555** | .487 | **.907** | .881 | .894 | .558 | **.576** | .563 | | |
| Reddit | Naïve | Majority | – | .000 | .000 | – | .000 | .000 | .724 | 1.000 | .840 | – | .333 | .280 | | |
| | | Random | .066 | .066 | .066 | .158 | .158 | .158 | .776 | .776 | .776 | .333 | .333 | .333 | | |
| | | IIITH (Boinepelli et al., 2022) | .206 | **.524** | .296 | .402 | .630 | .491 | **.954** | .647 | .771 | .520 | .600 | .519 | | |
| | Timeline-level | LAMA (AlHamed et al., 2022) | .166 | .354 | .226 | .609 | .389 | .475 | .882 | .861 | .871 | .552 | .535 | .524 | | |
| | (CLPsych) | WResearch (Bayram and Benhiba, 2022) | .362 | .256 | .300 | .646 | .553 | .596 | .868 | .929 | .897 | .625 | .579 | .598 | ✓ | |
| | | UoS (Azim et al., 2022) | **.490** | .305 | .376 | **.697** | .630 | **.662** | .881 | **.940** | **.909** | **.689** | .625 | .649 | ✓ | ✓ |
| | Timeline-level | SBERT(avg hist) | .340 | .329 | .330 | .605 | .563 | .582 | .893 | .912 | .902 | .613 | .601 | .605 | | |
| | (-signature) | BiLSTM-sbert(hist) | .463 | .407 | **.430** | .629 | **.637** | .630 | .895 | .901 | .898 | .663 | **.648** | **.653** | | |
| | Timeline-level (+signature) | Seq-Sig-Net (our work) | .454 | .405 | .425 | .643 | .607 | .624 | .896 | .919 | .908 | .664 | .644 | .652 | | |

# Results

## real-time application

| | | | IS | | | IE | | | O | | | macro-avg | | | Model Type | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 | Emotion | Future |
| TalkLife | Naïve | Majority | – | – | – | – | – | – | .845 | 1 | .916 | .282 | .333 | .305 | | |
| | | Random | .047 | .047 | .047 | .108 | .108 | .108 | .845 | .845 | .845 | .333 | .333 | .333 | | |
| | Post-level | BERT(f) (Tsakalidis et al., 2022b) | .260 | **.321** | .287 | .401 | .478 | .436 | .898 | .864 | .881 | .520 | .554 | .534 | | |
| | Timeline-level (Tsakalidis et al., 2022b) | EM-DM | **.553** | .118 | .193 | .479 | .351 | .405 | .880 | **.948** | .913 | **.631** | .472 | .504 | ✓ | ✓ |
| | | BiLSTM-bert | .397 | .264 | **.316** | **.568** | .461 | **.508** | .898 | .936 | **.917** | .621 | .553 | **.580** | | ✓ |
| | Timeline-level (-signature) | SBERT(avg hist) | .283 | .244 | .262 | .424 | .486 | .452 | .896 | .885 | .890 | .534 | .539 | .535 | | |
| | | BiLSTM-sbert(hist) | .258 | .272 | .264 | .442 | .506 | .468 | .901 | .879 | .890 | .534 | .553 | .541 | | |
| | | BiLSTM-bert(hist) | .405 | .241 | .302 | .536 | .415 | .468 | .892 | .938 | .914 | .611 | .531 | .561 | | |
| | Timeline-level (+signature) | Seq-Sig-Net (our work) | .331 | .290 | .309 | .435 | **.555** | .487 | **.907** | .881 | .894 | .558 | **.576** | .563 | | |
| Reddit | Naïve | Majority | – | .000 | .000 | – | .000 | .000 | .724 | 1.000 | .840 | – | .333 | .280 | | |
| | | Random | .066 | .066 | .066 | .158 | .158 | .158 | .776 | .776 | .776 | .333 | .333 | .333 | | |
| | Timeline-level (CLPsych) | IIITH (Boinepelli et al., 2022) | .206 | **.524** | .296 | .402 | .630 | .491 | **.954** | .647 | .771 | .520 | .600 | .519 | | |
| | | LAMA (AlHamed et al., 2022) | .166 | .354 | .226 | .609 | .389 | .475 | .882 | .861 | .871 | .552 | .535 | .524 | | |
| | | WResearch (Bayram and Benhiba, 2022) | .362 | .256 | .300 | .646 | .553 | .596 | .868 | .929 | .897 | .625 | .579 | .598 | ✓ | |
| | | UoS (Azim et al., 2022) | **.490** | .305 | .376 | **.697** | .630 | **.662** | .881 | **.940** | **.909** | **.689** | .625 | .649 | ✓ | ✓ |
| | Timeline-level (-signature) | SBERT(avg hist) | .340 | .329 | .330 | .605 | .563 | .582 | .893 | .912 | .902 | .613 | .601 | .605 | | |
| | | BiLSTM-sbert(hist) | .463 | .407 | **.430** | .629 | **.637** | .630 | .895 | .901 | .898 | .663 | **.648** | **.653** | | |
| | Timeline-level (+signature) | Seq-Sig-Net (our work) | .454 | .405 | .425 | .643 | .607 | .624 | .896 | .919 | .908 | .664 | .644 | .652 | | |

# Results

## real-time application

| | | | IS | | | IE | | | O | | | macro-avg | | | Model Type | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 | Emotion | Future |
| TalkLife | Naïve | Majority | – | – | – | – | – | – | .845 | 1 | .916 | .282 | .333 | .305 | | |
| | | Random | .047 | .047 | .047 | .108 | .108 | .108 | .845 | .845 | .845 | .333 | .333 | .333 | | |
| | Post-level | BERT(f) (Tsakalidis et al., 2022b) | .260 | **.321** | .287 | .401 | .478 | .436 | 898 | .864 | .881 | .520 | .554 | .534 | | |
| | Timeline-level | EM-DM | **.553** | .118 | .193 | .479 | .351 | .405 | .880 | **.948** | .913 | **.631** | .472 | .504 | ✓ | ✓ |
| | (Tsakalidis et al., 2022b) | BiLSTM-bert | .397 | .264 | **.316** | **.568** | .461 | **.508** | .898 | .936 | **.917** | .621 | .553 | **.580** | | ✓ |
| | Timeline-level | SBERT(avg hist) | .283 | .244 | .262 | .424 | .486 | .452 | .896 | .885 | .890 | .534 | .539 | .535 | | |
| | (-signature) | BiLSTM-sbert(hist) | .258 | .272 | .264 | .442 | .506 | .468 | .901 | .879 | .890 | .534 | .553 | .541 | | |
| | | BiLSTM-bert(hist) | .405 | .241 | .302 | .536 | .415 | .468 | .892 | .938 | .914 | .611 | .531 | .561 | | |
| | Timeline-level (+signature) | Seq-Sig-Net (our work) | .331 | .290 | .309 | .435 | **.555** | .487 | **.907** | .881 | .894 | .558 | **.576** | .563 | | |
| Reddit | Naïve | Majority | – | .000 | .000 | – | .000 | .000 | .724 | 1.000 | .840 | – | .333 | .280 | | |
| | | Random | .066 | .066 | .066 | .158 | .158 | .158 | .776 | .776 | .776 | .333 | .333 | .333 | | |
| | | IIITH (Boinepelli et al., 2022) | .206 | **.524** | .296 | .402 | .630 | .491 | **.954** | .647 | .771 | .520 | .600 | .519 | | |
| | Timeline-level | LAMA (AlHamed et al., 2022) | .166 | .354 | .226 | .609 | .389 | .475 | .882 | .861 | .871 | .552 | .535 | .524 | | |
| | (CLPsych) | WResearch (Bayram and Benhiba, 2022) | .362 | .256 | .300 | .646 | .553 | .596 | .868 | .929 | .897 | .625 | .579 | .598 | ✓ | |
| | | UoS (Azim et al., 2022) | **.490** | .305 | .376 | **.697** | .630 | **.662** | .881 | **.940** | **.909** | **.689** | .625 | .649 | ✓ | ✓ |
| | Timeline-level | SBERT(avg hist) | .340 | .329 | .330 | .605 | .563 | .582 | .893 | .912 | .902 | .613 | .601 | .605 | | |
| | (-signature) | BiLSTM-sbert(hist) | .463 | .407 | **.430** | .629 | **.637** | .630 | .895 | .901 | .898 | .663 | **.648** | **.653** | | |
| | Timeline-level (+signature) | Seq-Sig-Net (our work) | .454 | .405 | .425 | .643 | .607 | .624 | .896 | .919 | .908 | .664 | .644 | .652 | | |

# Results

## real-time application

## generalisable

| | | | IS | | | IE | | | O | | | macro-avg | | | Model Type | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | P | R | F1 | P | R | F1 | P | R | F1 | P | R | F1 | Emotion | Future |
| TalkLife | Naïve | Majority | – | – | – | – | – | – | .845 | 1 | .916 | .282 | .333 | .305 | | |
| | | Random | .047 | .047 | .047 | .108 | .108 | .108 | .845 | .845 | .845 | .333 | .333 | .333 | | |
| | Post-level | BERT(f) (Tsakalidis et al., 2022b) | .260 | **.321** | .287 | .401 | .478 | .436 | 898 | .864 | .881 | .520 | .554 | .534 | | |
| | Timeline-level (Tsakalidis et al., 2022b) | EM-DM | **.553** | .118 | .193 | .479 | .351 | .405 | .880 | **.948** | .913 | **.631** | .472 | .504 | ✓ | ✓ |
| | | BiLSTM-bert | .397 | .264 | **.316** | **.568** | .461 | **.508** | .898 | .936 | **.917** | .621 | .553 | **.580** | | ✓ |
| | Timeline-level (-signature) | SBERT(avg hist) | .283 | .244 | .262 | .424 | .486 | .452 | .896 | .885 | .890 | .534 | .539 | .535 | | |
| | | BiLSTM-sbert(hist) | .258 | .272 | .264 | .442 | .506 | .468 | .901 | .879 | .890 | .534 | .553 | .541 | | |
| | | BiLSTM-bert(hist) | .405 | .241 | .302 | .536 | .415 | .468 | .892 | .938 | .914 | .611 | .531 | .561 | | |
| | Timeline-level (+signature) | Seq-Sig-Net (our work) | .331 | .290 | .309 | .435 | **.555** | .487 | **.907** | .881 | .894 | .558 | **.576** | .563 | | |
| Reddit | Naïve | Majority | – | .000 | .000 | – | .000 | .000 | .724 | 1.000 | .840 | – | .333 | .280 | | |
| | | Random | .066 | .066 | .066 | .158 | .158 | .158 | .776 | .776 | .776 | .333 | .333 | .333 | | |
| | Timeline-level (CLPsych) | IIITH (Boinepelli et al., 2022) | .206 | **.524** | .296 | .402 | .630 | .491 | **.954** | .647 | .771 | .520 | .600 | .519 | | |
| | | LAMA (AlHamed et al., 2022) | .166 | .354 | .226 | .609 | .389 | .475 | .882 | .861 | .871 | .552 | .535 | .524 | | |
| | | WResearch (Bayram and Benhiba, 2022) | .362 | .256 | .300 | .646 | .553 | .596 | .868 | .929 | .897 | .625 | .579 | .598 | ✓ | |
| | | UoS (Azim et al., 2022) | **.490** | .305 | .376 | **.697** | .630 | **.662** | .881 | **.940** | **.909** | **.689** | .625 | .649 | ✓ | ✓ |
| | Timeline-level (-signature) | SBERT(avg hist) | .340 | .329 | .330 | .605 | .563 | .582 | .893 | .912 | .902 | .613 | .601 | .605 | | |
| | | BiLSTM-sbert(hist) | .463 | .407 | **.430** | .629 | **.637** | .630 | .895 | .901 | .898 | .663 | **.648** | **.653** | | |
| | Timeline-level (+signature) | Seq-Sig-Net (our work) | .454 | .405 | .425 | .643 | .607 | .624 | .896 | .919 | .908 | .664 | .644 | .652 | | |

# Computational Resources

Seq-Sig-Net is much greener

| Model name | Memory (MB) | Parameters (million) | Avg Training time (minutes) |
|---|---|---|---|
| BiLSTM-bert(hist) | 18.9 | 2.5 | 36.7 |
| Seq-Sig-Net | 12.9 | 1.7 | 33.9 |

# Ablation Study

| | | TalkLife | | | | Reddit | | | |
|---|---|---|---|---|---|---|---|---|---|
| Model name | Explanation of ablation | IS | IE | O | avg | IS | IE | O | avg |
| SBERT post | (*) | .281 | .431 | .887 | .533 | .200 | .541 | .909 | .550 |
| SBERT(avg hist) | (*) + mean hist. + t | .262 | .452 | .890 | .535 | .330 | .582 | .902 | .605 |
| SWNU Network | (*) + 1 SWNU + t | .296 | .477 | **.894** | .556 | .308 | .623 | **.911** | .614 |
| Seq-Sig-Net | (*) + BiLSTM on SWNU + t | **.309** | **.487** | **.894** | **.563** | **.425** | **.624** | .908 | **.652** |

# Ablation Study

## efficient time windows

| | | TalkLife | | | | Reddit | | | |
|---|---|---|---|---|---|---|---|---|---|
| Model name | Explanation of ablation | IS | IE | O | avg | IS | IE | O | avg |
| SBERT post | (*) | .281 | .431 | .887 | .533 | .200 | .541 | .909 | .550 |
| SBERT(avg hist) | (*) + mean hist. + t | .262 | .452 | .890 | .535 | .330 | .582 | .902 | .605 |
| SWNU Network | (*) + 1 SWNU + t | .296 | .477 | **.894** | .556 | .308 | .623 | **.911** | .614 |
| Seq-Sig-Net | (*) + BiLSTM on SWNU + t | **.309** | **.487** | **.894** | **.563** | **.425** | **.624** | .908 | **.652** |

# Ablation Study

## efficient time windows

## memorising local parts of timeline

| | | TalkLife | | | | Reddit | | | |
|---|---|---|---|---|---|---|---|---|---|
| Model name | Explanation of ablation | IS | IE | O | avg | IS | IE | O | avg |
| SBERT post | (*) | .281 | .431 | .887 | .533 | .200 | .541 | .909 | .550 |
| SBERT(avg hist) | (*) + mean hist. + t | .262 | .452 | .890 | .535 | .330 | .582 | .902 | .605 |
| SWNU Network | (*) + 1 SWNU + t | .296 | .477 | **.894** | .556 | .308 | .623 | **.911** | .614 |
| Seq-Sig-Net | (*) + BiLSTM on SWNU + t | **.309** | **.487** | **.894** | **.563** | **.425** | **.624** | .908 | **.652** |

# Analysis

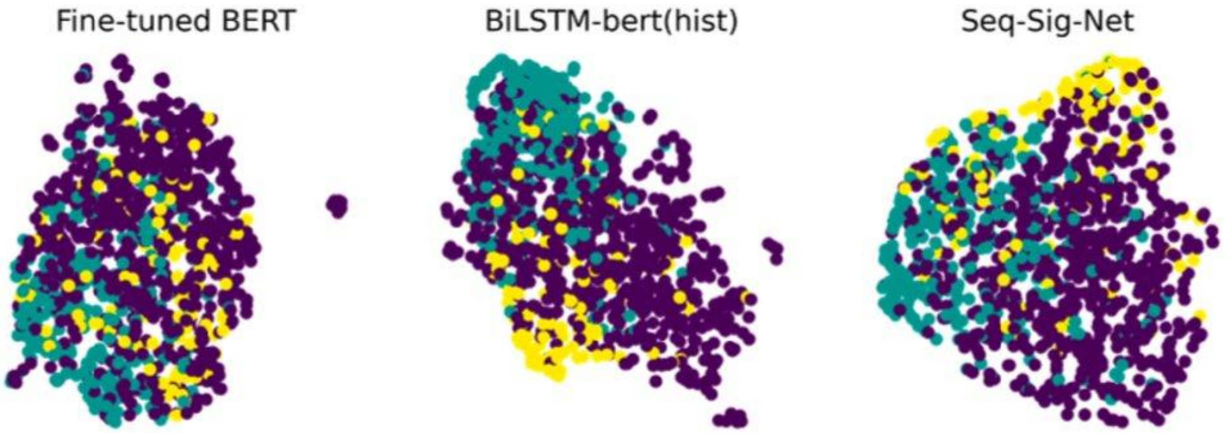
Clustering ability of representations?

# Analysis

## Clustering ability of representations?

**Sequentiality - better**

| | Silhouette (~1) | Calinski Harabasz ↑ | Davies Bouldin ↓ |
|---|---|---|---|
| BERT fine-tuned | -0.091 | 134.01 | 3.15 |
| BiLSTM-bert(hist) | -0.050 | 275.51 | 2.59 |
| Seq-Sig-Net | **-0.014** | **294.66** | **2.45** |



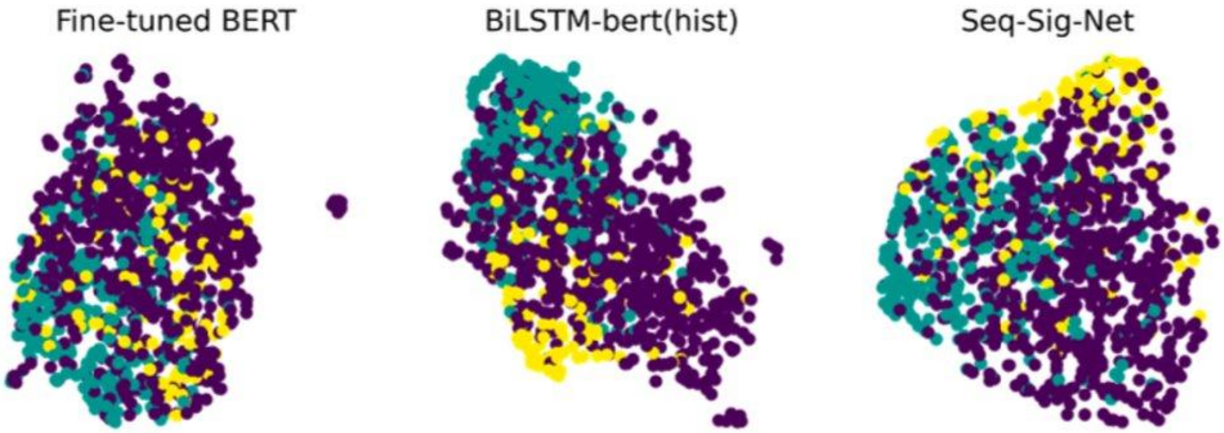
# Analysis

**Clustering ability of representations?**

**Sequentiality - better**

**Signatures - even better**

| | Silhouette (~1) | Calinski Harabasz ↑ | Davies Bouldin ↓ |
|---|---|---|---|
| BERT fine-tuned | -0.091 | 134.01 | 3.15 |
| BiLSTM-bert(hist) | -0.050 | 275.51 | 2.59 |
| Seq-Sig-Net | **-0.014** | **294.66** | **2.45** |



# Contributions

**Signature Transforms in Neural Networks for Language Modeling.**

# Contributions

**Signature Transforms** in Neural Networks for Language Modeling.

Generalisable to sequential **real-time** applications.

# Contributions

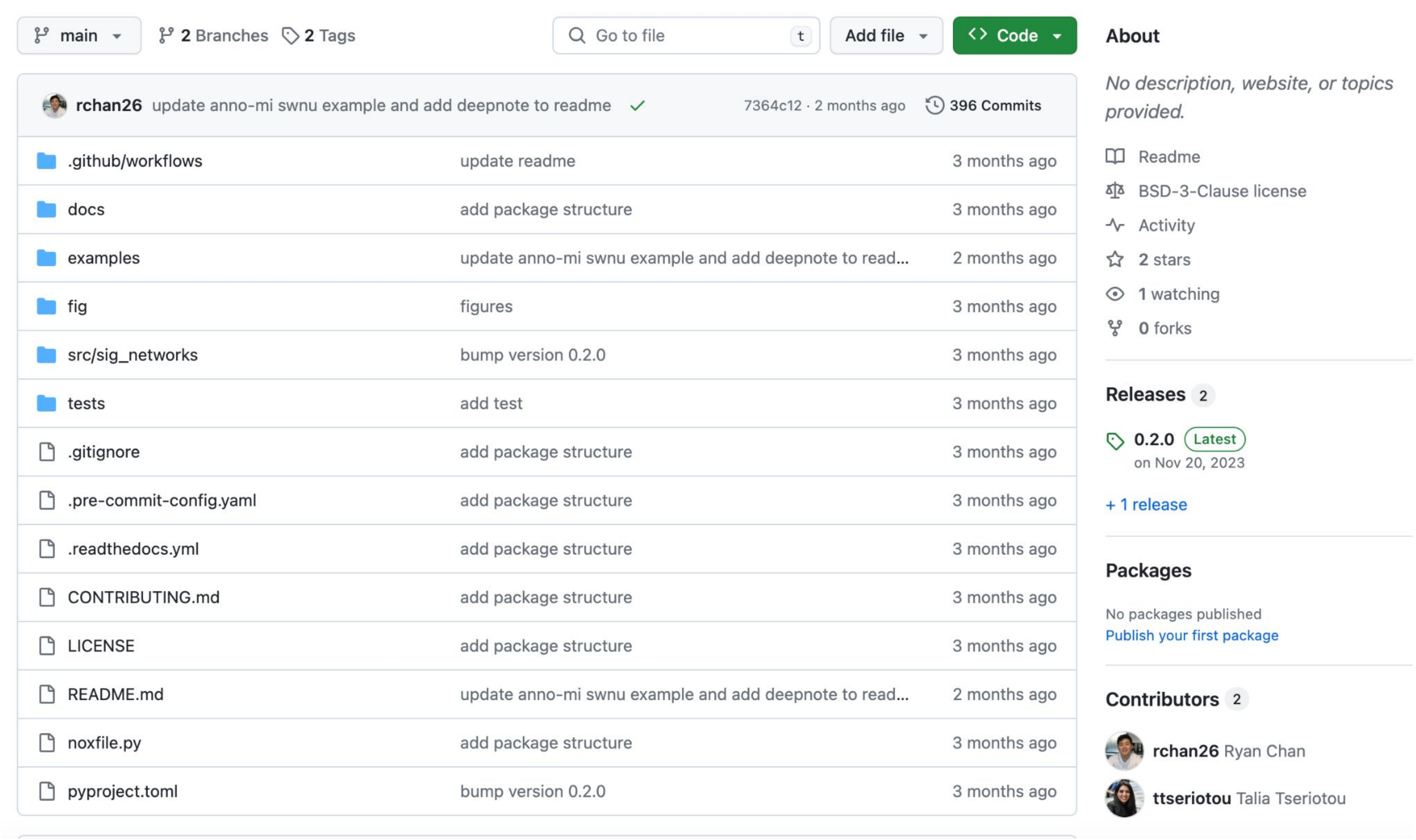
**Signature Transforms in Neural Networks for Language Modeling.**

**Generalisable to sequential real-time applications.**

**SOTA performance against historical user data baselines.**

# Sig-Networks Toolkit

**pip installable PyTorch package for longitudinal NLP modelling**.




https://github.com/ttseriotou/sig-networks

# Contributions

1. **sig-networks** open-source pip installable toolkit for longitudinal NLP tasks.

# Contributions

1. **sig-networks** open-source pip installable toolkit for longitudinal NLP tasks.
2. **nlpsig** pip installable library for data preprocessing.

# Contributions

1. **sig-networks** open-source pip installable toolkit for longitudinal NLP tasks.
2. **nlpsig** pip installable library for data preprocessing.
3. **SOTA** performance on three tasks.

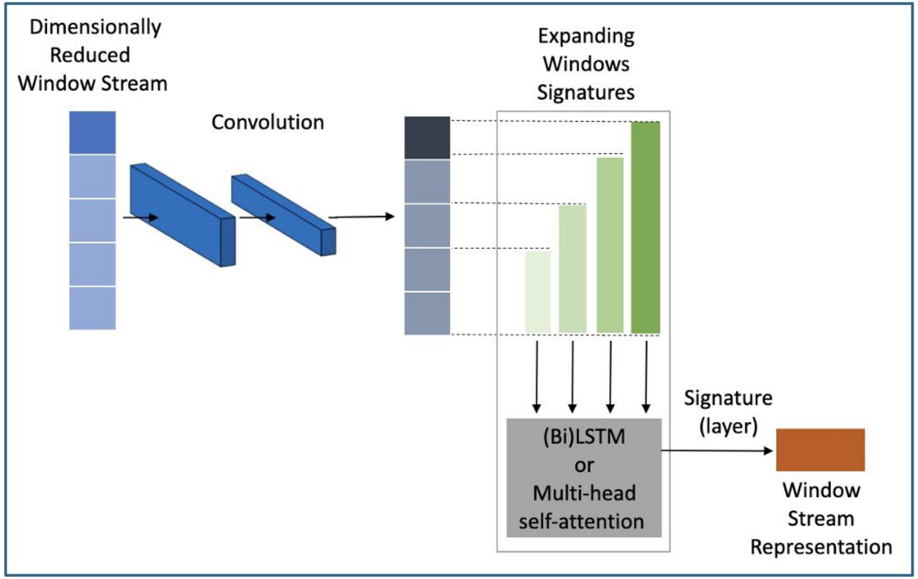
# Contributions

1. **sig-networks** open-source pip installable toolkit for longitudinal NLP tasks.
2. **nlpsig** pip installable library for data preprocessing.
3. **SOTA** performance on three tasks.
4. Flexible dataset adaptation, model building blocks, benchmarking, feature and parameter selection.

# Signature Network Models

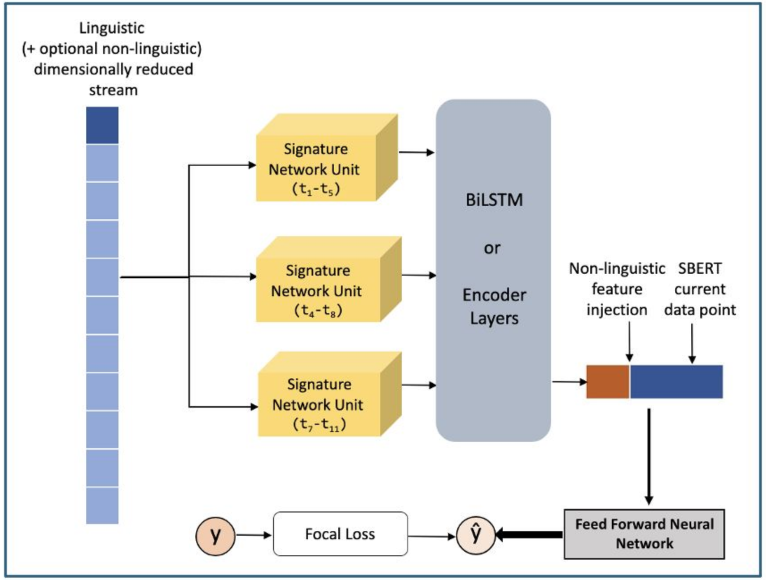
**Window-based**

Signatures over short expanding windows fed in BiLSTM/MHA

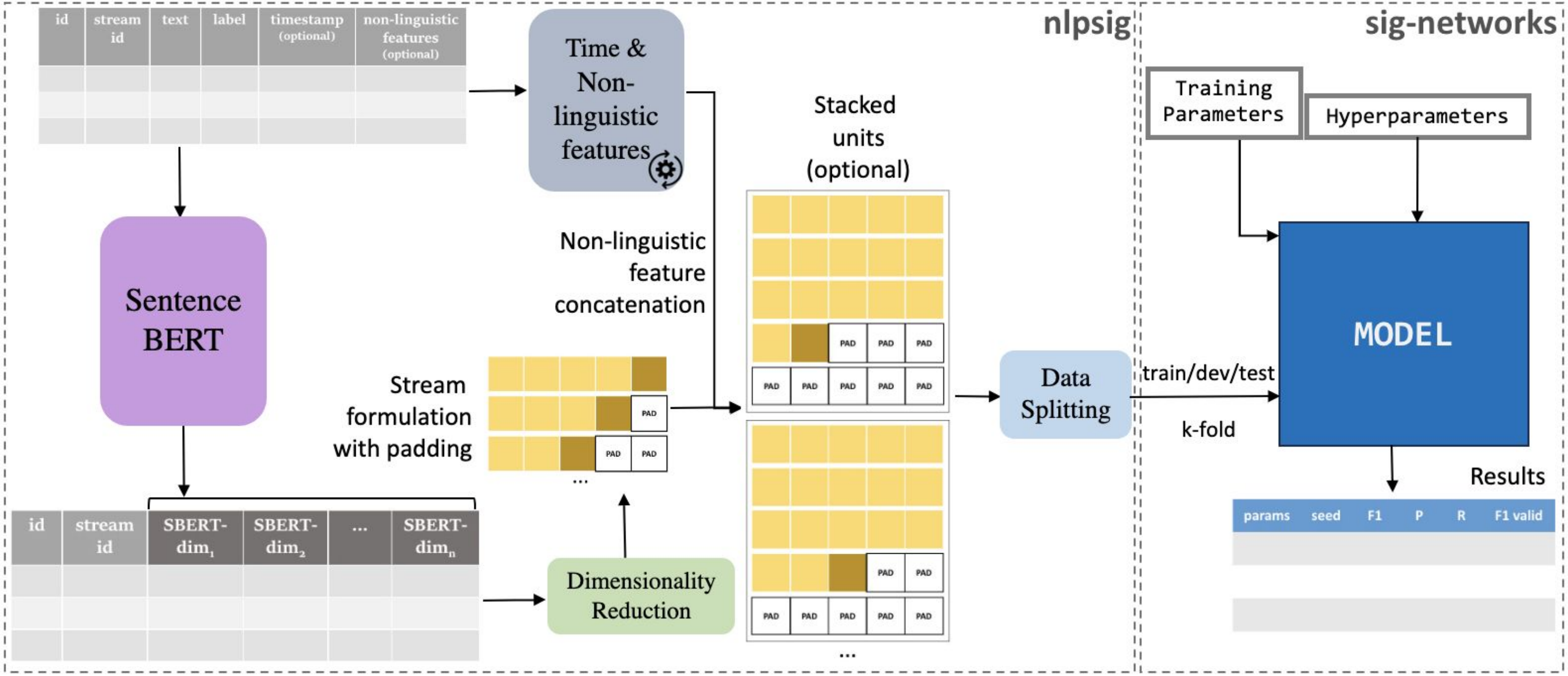



**Unit-based**

Sequential modeling of window units through a BiLSTM/Encoder



# System Overview



# System Components

## Data Preparation and Training Modules

- nlpsig.encode_text → embedding generation
- nlpsig.DimReduce →dimensionality reduction
- nlpsig.PrepareData →padded embedding streams, time features calculation
- nlpsig.classification_utils → k-fold cross validation, stratified splitting, user-defined folds, data points exclusion.
- User selections: loss function, validation metric, patience, random seeds, grid search.

# System Components

### Data Preparation and Training Modules

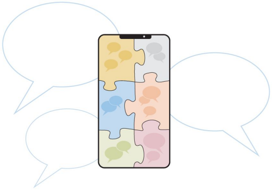
- nlpsig.encode_text → embedding generation
- nlpsig.DimReduce →dimensionality reduction
- nlpsig.PrepareData →padded embedding streams, time features calculation
- nlpsig.classification_utils → k-fold cross validation, stratified splitting, user-defined folds, data points exclusion.
- User selections: loss function, validation metric, patience, random seeds, grid search, feature concatenation in/out of path

### Model Modules

- PyTorch classes as building blocks of our models to allow new systems development.
- Baselines: BERT, FFN (with/out history stream), BiLSTM
- Signature Network Models: range of options i.e. no. encoder layers.

# Data

| | AnnoMI | Longitudinal Rumour Stance | TalkLife |
|---|---|---|---|
| Description | Counselling Dialogues | Twitter conversations discussing rumours | Posts from peer-to-peer support network |
| Timelines | 133 | 325 | 500 |
| Data points | 4,817 (client utterances) | 5,568 | 18,604 |
| Predicting | Client response type | Switch in the aggregate stance towards claim | Change in user's mood |



# Results

To appear in EACL 2024 Sig-Networks Toolkit: Signature Networks for Longitudinal Language Modelling

| Model | Anno-MI (3-class) | | | LRS (2-class) | | | TalkLife (3-class) | | |
|---|---|---|---|---|---|---|---|---|---|
| BERT (focal) | .519 | | | .589 | | | .531 | | |
| BERT (ce) | .501 | | | .596 | | | .521 | | |
| FFN | .512 | | | .581 | | | .534 | | |
| FFN History | .520 | | | .625 | | | .537 | | |
| BiLSTM ($w = 5$) | .517 | | | .637 | | | .544 | | |
| SWNU ($w = 5$) | .522 | | | .670 | | | **.563** | | |
| SW-Attn ($w = 5$) | .515 | | | .667 | | | .556 | | |
| **History Length**<br>**#units** (**$w$=5, $k$=3**) | **11**<br>**3** | **20**<br>**6** | **35**<br>**11** | **11**<br>**3** | **20**<br>**6** | **35**<br>**11** | **11**<br>**3** | **20**<br>**6** | **35**<br>**11** |
| BiLSTM | .518 | .507 | .510 | .657 | .648 | .648 | .539 | .533 | .525 |
| SWNU | .522 | .512 | .493 | .671 | .654 | <u>.673</u> | .550 | .537 | .539 |
| SW-Attn | .517 | .508 | .508 | .659 | .665 | .661 | .547 | .541 | .539 |
| Seq-Sig-Net | **.525** | <u>.523</u> | .517 | .672 | **.678** | .654 | **.563** | <u>.561</u> | .559 |
| SW-Attn+BiLSTM | .511 | .514 | .515 | .663 | .657 | .660 | .554 | .557 | .550 |
| SW-Attn+Encoder | .498 | .506 | .505 | .664 | .657 | .662 | .552 | .552 | .545 |

# Results

- Seq-Sig-Net achieves SOTA or on-par with SWNU across all tasks

| Model | Anno-MI (3-class) | | | LRS (2-class) | | | TalkLife (3-class) | | |
|---|---|---|---|---|---|---|---|---|---|
| BERT (focal) | .519 | | | .589 | | | .531 | | |
| BERT (ce) | .501 | | | .596 | | | .521 | | |
| FFN | .512 | | | .581 | | | .534 | | |
| FFN History | .520 | | | .625 | | | .537 | | |
| BiLSTM ($w = 5$) | .517 | | | .637 | | | .544 | | |
| SWNU ($w = 5$) | .522 | | | .670 | | | **.563** | | |
| SW-Attn ($w = 5$) | .515 | | | .667 | | | .556 | | |
| **History Length** | **11** | **20** | **35** | **11** | **20** | **35** | **11** | **20** | **35** |
| **#units ($w$=5, $k$=3)** | **3** | **6** | **11** | **3** | **6** | **11** | **3** | **6** | **11** |
| BiLSTM | .518 | .507 | .510 | .657 | .648 | .648 | .539 | .533 | .525 |
| SWNU | .522 | .512 | .493 | .671 | .654 | <u>.673</u> | .550 | .537 | .539 |
| SW-Attn | .517 | .508 | .508 | .659 | .665 | .661 | .547 | .541 | .539 |
| Seq-Sig-Net | **.525** | <u>.523</u> | .517 | .672 | **.678** | .654 | **.563** | <u>.561</u> | .559 |
| SW-Attn+BiLSTM | .511 | .514 | .515 | .663 | .657 | .660 | .554 | .557 | .550 |
| SW-Attn+Encoder | .498 | .506 | .505 | .664 | .657 | .662 | .552 | .552 | .545 |

# Results

- In LRS and TalkLife SigNetworks outperforms all baselines, for each history length
- Anno-MI is the least longitudinal - small performance gains of sequential models

| Model | Anno-MI (3-class) | | | LRS (2-class) | | | TalkLife (3-class) | | |
|---|---|---|---|---|---|---|---|---|---|
| BERT (focal) | .519 | | | .589 | | | .531 | | |
| BERT (ce) | .501 | | | .596 | | | .521 | | |
| FFN | .512 | | | .581 | | | .534 | | |
| FFN History | .520 | | | .625 | | | .537 | | |
| BiLSTM ($w = 5$) | .517 | | | .637 | | | .544 | | |
| SWNU ($w = 5$) | .522 | | | .670 | | | **.563** | | |
| SW-Attn ($w = 5$) | .515 | | | .667 | | | .556 | | |
| **History Length** | **11** | **20** | **35** | **11** | **20** | **35** | **11** | **20** | **35** |
| **#units ($w$=5, $k$=3)** | **3** | **6** | **11** | **3** | **6** | **11** | **3** | **6** | **11** |
| BiLSTM | .518 | .507 | .510 | .657 | .648 | .648 | .539 | .533 | .525 |
| SWNU | .522 | .512 | .493 | .671 | .654 | <u>.673</u> | .550 | .537 | .539 |
| SW-Attn | .517 | .508 | .508 | .659 | .665 | .661 | .547 | .541 | .539 |
| Seq-Sig-Net | **.525** | <u>.523</u> | .517 | .672 | **.678** | .654 | **.563** | <u>.561</u> | .559 |
| SW-Attn+BiLSTM | .511 | .514 | .515 | .663 | .657 | .660 | .554 | .557 | .550 |
| SW-Attn+Encoder | .498 | .506 | .505 | .664 | .657 | .662 | .552 | .552 | .545 |

# Time-Scale Analysis

- Different degree of temporal granularity

| Dataset | Anno-MI | | Longitudinal Rumour Stance | TalkLife MoC | |
|---|---|---|---|---|---|
| | Change | Sustain | Switch | Switch | Escalation |
| Mean Point Time Diff. | 5sec | | 1hr 26min 40sec | 6hr 51min 11sec | |
| Median Point Time Diff. | 3sec | | 1min 39sec | 59min 38sec | |
| Mean consecutive events | 2.21 | 1.68 | 8.52 | 1.58 | 4.12 |
| Median consecutive events | 1 | 1 | 4 | 1 | 3 |
| Mean no. of events in stream | 8.86 | 4.05 | 6.45 | 1.77 | 4.03 |
| Median no. of events in stream | 5 | 3 | 0 | 1 | 1 |

# What’s next?

Moving away from pre-trained representations …

# Objectives

**O1** Transformer extension to meaningfully incorporate **temporal relationship information**

**O2** Applicability of temporal Transformer to **longitudinal tasks**

**O3** Ability to incorporate information from **long temporally distanced** text over time periods of different granularity

# Contributions

1. **SOTA** performance on three different longitudinal tasks.

# Contributions

1. **SOTA** performance on three different longitudinal tasks.
2. Rotary **temporal** positional embeddings

   → measure temporal distance between points.

# Contributions

1. **SOTA** performance on three different longitudinal tasks.
2. Rotary **temporal** positional embeddings

   $\longrightarrow$ measure temporal distance between points.
3. Simultaneous fine-tuning of word embeddings and temporal data streams through a sequential transformer layer

   $\longrightarrow$ domain and longitudinal model adaptation.

# Contributions

1. **SOTA** performance on three different longitudinal tasks.
2. Rotary **temporal** positional embeddings

   → measure temporal distance between points.
3. Simultaneous fine-tuning of word embeddings and temporal data streams through a sequential transformer layer

   → domain and longitudinal model adaptation.
4. Sequential transformer layer is flexible in terms of layer and architecture.

*The rest of the slides were removed as they contain unpublished work*

Thank you!