Annotating Moments of Change (Change-points): Guidelines

Goal

The goal of the Turing AI Fellowship is to create temporally sensitive models for capturing changes in individuals over time by leveraging language and other digital content they produce, which we call user generated content (UGC). The methods we develop consist in representing users & their behaviour over time through their UGC, tracking changes in these representations and summarising such changes and the reasons behind them. We are particularly interested in capturing *moments of change*. The ultimate goal is to be able to use our methods for automated detection of moments of change in creating new tools to monitor mental health and the progression of individuals.

Definition of moment of change

Change in an individual's behaviour or mood after a particular event or the identifiable start of a trend. A moment of change may be a particular point in time or a longer period. It can be an obvious change or a more subtle change in one's mood. Moments of change can be defined differently in different settings (e.g. different mental health conditions or clinical settings), especially when annotating posts over a long period of time (e.g. posts from an online platform) as opposed to within a therapy session or interview with a carer or clinician.

(A) Types of moments of change over a longer period

These moments of change are particularly applicable to platform data and include**:**

1. Change in behaviour or mood. This can include the start of a period of low mood or the escalation of a previously identified mood (e.g. starting to discuss suicidality when previous posts were about depression or standard topics). It can also be a switch (e.g. starting to make positive comments after having a low mood previously or demonstrating a low mood after neutral or positive posts).
2. An important life event (e.g. death of a loved one, a romantic breakup), even if we can't see the immediate effects in the content shared by the user afterwards.
3. The occurrence of a physical symptom (e.g. pain, lack of sleep etc.).

(B) Types of moments of change over a single conversation

The above in A (especially A(1)) is also applicable in this setting. Additionally we can have:

1. "Aha" moments: Moments of positive insight or improvement in a conversation.
2. Difficult moments in the conversation: the conversation pauses, there is a potential rupture between the individual and the carer or participant.
3. Moments of intervention: the carer/clinician makes a particular suggestion or changes topic.

Defining Timelines

An individual's timeline is a series of actions and interactions over a particular time period. Such "actions" can be the utterances of an individual during conversations with his/her therapist or carer, or the content of an individual's posts and other digital interactions on an online platform, within the specified time period.

For the purposes of our current annotation scheme, we focus on timelines created from individuals' posts on an online platform, where users discuss issues around their mental health state and related thoughts. In particular, *we consider the posts shared by a single person over a particular time period as their "timeline"* (see "Timeline Generation" below for technical details on how such time periods are being identified). The posts within a timeline are presented on a daily basis, along with their respective comments and meta-data (i.e., associated mood tagged by the user on each post-where available, the time that the post was shared). Examples from anonymised individuals' timelines are provided in this document, along with screenshots of the annotation interface and associated annotation tasks.

Timeline Generation: We generate univariate time-series of a variable X for a single individual based on posts they have shared and their interactions with other online users. Based on the time series of the variable X, we try to find anomalies regarding its value. These potential anomalies will serve as anchors for the timeline. Depending on the variable X, we generate different daily time series for:

- Posting behaviour (number of posts over time);
- Commenting (number of comments received over time);

For each time series and for each individual independently, we detect anomalous periods of activities via different approaches. These include (a) change-point detection algorithms, (b) motif-based discoveries and (c) detection of periods containing one or more days of unexpected (probability <1%) and extremely high/low values of the variable under study (posts/comments) and extracting a certain time window around those.

Annotation Tasks

A1a. Mood (Switch/Escalation)

Given the timeline of an individual generated as above, we ask annotators to read through the posts shared by the individual on each day included within the timeline and **annotate those posts that denote a change in the user's mood,** compared to the recent past.

Note that there could be two types of mood changes: drastic (a switch) or gradual (an escalation).

In particular we are interested in cases where there is an <u>intensification</u> in one's mood whether it be positive or negative (**escalation**). We are also looking for a <u>drastic</u> change of one's mood from relatively neutral or positive to negative and vice versa (**switch**). Escalations and switches can take place over several posts or days or within a single post. Therefore we are interested in the range of posts denoting a particular switch or escalation.

Take away:

There are a few important things to remember when annotating switches and escalations:

1. A **switch** is a **drastic** change in mood, in comparison with the recent past. We annotate the <u>first</u> post (which should have a clearly different mood in comparison with the previous one) and we add as additional information the following range (how many other posts are also related to this switch)

2. An **escalation** is a **gradual** change in mood, which should last for a few posts. In this case, we annotate what we consider <u>the main post</u> (the peak of the escalation) and add as additional information the preceding and following range (how many other posts are part of this gradual change in mood)

Consider the example in (1) below. Here the user had been in a low mood with feelings of procrastination. Then they switch to more positive feelings. At this point we capture the <u>first post</u> where the switch starts and denote the range of posts signalling the switch, which in this case are 6.1 and 6.2.

(1)

Tuesday, 26 Jan 2016
4.1.I've been feeling lonely for a long time

☐

SHOW/HIDE CONVERSATIONS

Wednesday, 27 Jan 2016
5.1.been procrastinating so much these days

☐

SHOW/HIDE CONVERSATIONS

Ø

Friday, 29 Jan 2016
6.1.today is finally a new day, i'm getting some work done and it's only 10am!!
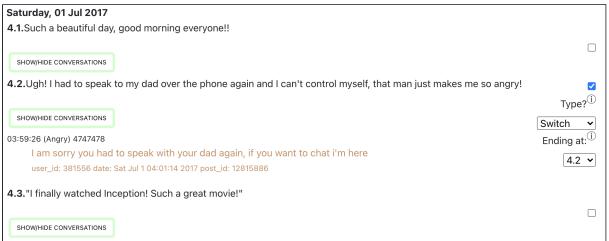
☑

Type?ⓘ

SHOW/HIDE CONVERSATIONS

Switch ▾

Ending at:ⓘ

6.2 ▾

6.2.yes it really feels different, so happy finally

☐

SHOW/HIDE CONVERSATIONS

In example (2) below, the user seems to be in a good mood. Then in the next post (a few hours later) they get angry by talking to their father. This marks a switch at post 4.2, so we select "switch" as the type and add 4.2 in the post range. If the anger had lasted several more posts we would have added the corresponding posts in the "Ending at" range.

(2)

Saturday, 01 Jul 2017
4.1. Such a beautiful day, good morning everyone!!

☐

SHOW/HIDE CONVERSATIONS

4.2. Ugh! I had to speak to my dad over the phone again and I can't control myself, that man just makes me so angry!

☑

Type?ⓘ

SHOW/HIDE CONVERSATIONS

Switch ▾

03:59:26 (Angry) 4747478

Ending at:ⓘ

I am sorry you had to speak with your dad again, if you want to chat i'm here

4.2 ▾

user_id: 381556 date: Sat Jul 1 04:01:14 2017 post_id: 12815886

4.3."I finally watched Inception! Such a great movie!"

☐

SHOW/HIDE CONVERSATIONS

In example (3) we have a switch where the user goes from feeling bored to feeling excited in 4.4. Then from 4.5 we have an escalation with a peak at 4.7 where the user says they feel extremely sad. The post range for the escalation is 4.4-4.9.

(3)

4.2. Nothing to do

☐

METADATA

4.3. Im bored

☐

METADATA

4.4.downloading some cool new games, finally!

☑

Type?ⓘ

METADATA

Switch ▼

Ending at:ⓘ

4.4 ▼

4.5.I feel so low can't sleep

☐

METADATA

4.6.I hate myself so much

☐

SHOW/HIDE CONVERSATIONS

4.7.I don't deserve happiness

☑

Type?ⓘ

SHOW/HIDE CONVERSATIONS

Escalation ▼

Starting at:ⓘ

4.5 ▼

Ending at:ⓘ

4.9 ▼

4.8.smoking too much these days

☐

METADATA

4.9. Listening to Chopin cause I got nothing better to do.

In example (4) below we have another case of escalation. The escalation starts in post 7.1, where the user says they feel bored and culminates with them hating themselves (8.3) and having suicidal ideation in post 8.5. In this case we would annotate the escalation at 8.5 (the peak) and provide the post range as 7.1-8.5.

(4)

Saturday, 07 Apr 2018
7.1.feel like my life is getting boring

☐

SHOW/HIDE CONVERSATIONS

7.2.i really hate this

☐

SHOW/HIDE CONVERSATIONS

7.3.all i can do is dream of being a normal man 💔having a normal life

☐

METADATA

7.4. Wishing i was somebody else

☐

METADATA

7.5. Uugh

☐

METADATA

7.6.starting to hate being myself

☐

SHOW/HIDE CONVERSATIONS

Sunday, 08 Apr 2018
8.1. Life🤮

☐

SHOW/HIDE CONVERSATIONS

8.2.I wish i could be a emotionless monster

☐

SHOW/HIDE CONVERSATIONS

8.3.its 3 in the morning and i'm hating myself wishing i was not here

☐

SHOW/HIDE CONVERSATIONS

8.4.bout to go crazy

☐

SHOW/HIDE CONVERSATIONS

8.5.any painless way to commit suicide?

☑

SHOW/HIDE CONVERSATIONS

Type?ⓘ

Escalation ▾

Starting at:ⓘ

7.1 ▾

Ending at:ⓘ

8.5 ▾