Deep Learning Project Proposal

Tejit Pabari   Gitika Bose
tvp2107   gb2606

Minimal goal (imagine you're busier than expected with other classes, finals, work, etc). This is the minimal outcome you'll be happy with for your project in that scenario.

- A working model that can be used as a method of verification for the coders at the "Laboratory for Clinical and Developmental Studies" to increase their efficiency and help standardize the process of coding across researchers
- Coding is the process of segmenting and labelling passages from interviews done at the lab. The coding process tries to measure the level of internal and external detail provided by the participant when describing some future event. Since we have a copy of this manual, our minimal goal is to design a BERT model that codes a passage correctly according to the manual with some level of specificity. A summary of the kinds of labelling prescribed in the manual can be found on the next page.
- For our minimal product, we want to at least be able to "Isolate and define the imaged future event" and do "Text Segmentation and Primary Categorization" as described in the manual.
  - Determine the central event using the cue and event description
  - Identify each segment/detail
  - Distinguish between internal and external details

Expected goal (this is what you hope to accomplish if all goes as planned).

- A working model that can be used as a method of verification for the coders (and potentially replace the need for coding) at the "Laboratory for Clinical and Developmental Studies"
- Our expected goal is to design a BERT model that codes a passage correctly according to the manual with some more specificity. Once again, the summary of the kinds of labelling prescribed in the manual can be found on the next page. We expect to complete all three steps (including "Secondary Text Categorization")
- We will also try to predict whether the passage comes from a participant diagnosed with suicide ideation or not based on the counts of various labels and the word features for each label.

Stretch goal (if you have additional time to work on this, or you make progress faster than expected, this is what you hope to get done).

- After completing the expected goal (designing a working model that labels and classifies events), we will use either LIME or BERT to identify which features are most indicative for each sentence label and passage classification
- Alternatively, as an addendum to the expected goal, we will attempt to create some sort of BERT tutorial that explains the technology and its usage in further depth.

<u>**Coding Manual:**</u> <u>Summary</u>
<u>Isolate and define the imaged future event</u>
1. Identify the <u>central event</u> using the given description and the cue (the initial prompt given to the subject).
2. **Event:** An event is the following -
    a. **Time Restricted:** Anything within a day. If the subject doesn't talk about anything that comes under a day, then take the best restricted timespan.
    b. **Subject Present:** The subject must be present in the event
    c. **Singular:** It is not repeated
    d. **Sub-events, but connected, cohesive**: Sub-events can exist, however they should be connected and cohesive.
    e. **Most Detailed:** If there are more than 1 event, the most detailed event is the central event.

<u>Text Segmentation and Primary Categorization</u>
1. **Segment**: A segment is a standalone grammatical clause
    a. Provides a bit of information
    b. Can be part of a clause
    c. Contains "singular subject + predicate"
2. **Primary Text Categorization:** Text can be categorized into multiple primary ways. Of these Internal and External are the most important
    a. **Internal Details:** All events related to the central event. They consist of:
        i. What happens in the event
        ii. When the event happens
        iii. Where the event takes place
        iv. About what the subject is thinking
        v. About what the subject is feeling
        vi. About what the subject is perceiving
    b. **External:** All events not related to the central event. They typically consist of:
        i. Non-central events
        ii. General Semantic information
    c. **Repetition:** Lack of new information is marked as repetition and isn't counted
    d. **Corrective information:** If the subject corrects themselves, only the corrected information is counted.
    e. **Nonsense Words:** Any nonsense words are not counted
    f. **Speaking to the Experimenter:** Subject's comment on tasks are not counted
    g. **Repeating the cue:** If the subject reads the cue out loud, and it matches the central event, it is counted as Internal detail. Otherwise not counted
    h. **Cut off text:** Not counted

Distinguishing between internal and external details:
1. **Describing other people's thoughts, feelings and comments:**
   a. **Internal:** If the description reflects what the individual is thinking or feeling during the central event.
   b. **External:** If the description reflects thoughts the individual will have in general.
2. **Information about time:**
   a. **Internal:** If it provides information about the central event
   b. **External:** If it is general or semantic information and doesn't provide information about the central event
3. **List of possible scenarios (May/ Might):**
   a. **Internal:** If the possibilities are about the central event

Secondary Text Classification:
1. **Event Detail:** Describes the unfolding of the story.
   a. Happenings
   b. Who was there, weather, clothing, appearance
   c. Adjectives and adverbs that cant be assigned in another category.
   d. Note: PRIORITY given to more specific categories
2. **Place Detail:**
   a. localization in space, including countries, bodies of water, provinces, cities, streets, buildings, rooms, and locations within a room
   b. PERCEPTUAL: One's orientation
3. **Time Detail:**
   a. Life epoch, year, season, month, date, day of week, time of day, or clock time
   b. Information about when in time a particular event occurred
   c. PERCEPTUAL: Duration Information
4. **Perceptual:**
   a. Sensory: Auditory, Olfactory, tactile/pain, Taste, visual, spatial-temporal
   b. objects that are part of the perceptual landscape
   c. EVENT: Objects directly involved in unfolding of an event. If the subject is interacting with or possessing the object.
5. **Emotion/Thought Details:**
   a. pertains to the mental state of the subject, states, thoughts, opinions, expectations, or beliefs
   b. Inferences about subject's mental state
   c. "we" are thinking/feeling something
   d. EXTERNAL:
      i. Thoughts expressed in retrospect
      ii. Beliefs or opinions that are long-standing
   e. EVENT: Inferences about other people's mental state