

eRisk 2019: T3 results

No Author Given

No Institute Given

1 Task 3: Measuring the severity of the signs of depression

This is a new task in 2019. The task consists of estimating the level of depression from a thread of user submissions. For each user, the participants were given a history of postings and the participants had to fill a standard depression questionnaire (based on the evidence found in the history of postings).

The questionnaires are defined from Beck's Depression Inventory (BDI), which assesses the presence of feelings like sadness, pessimism, loss of energy, etc. The questionnaire contains 21 questions (see figs 1, 2).

The task aims at exploring the viability of automatically estimating the severity of multiple symptoms associated with depression. Given the user's history of writings, the algorithms had to estimate the user's response to each individual question. We collected questionnaires filled by Social Media users together with their history of writings (we extracted each history of writings right after the user provided us with the filled questionnaire). The questionnaires filled by the users (ground truth) were used to assess the quality of the responses provided by the participating systems.

The participants were given a dataset with multiple users (for each user, his history of writings is provided) and they were asked to produce a file with the following structure:

```
username1 answer1 answer2 .... answer21
username2 ....
....
```

Each line has an identification of the user and 21 values. These values correspond with the responses to the questions of the depression questionnaire (the possible values are 0, 1a, 1b, 2a, 2b, 3a, 3b -for questions 16 and 18- and 0, 1, 2, 3 -for the rest of the questions-).

2 Task 3: evaluation metrics

We considered a number of metrics in order to assess the quality of a questionnaire filled by a system when compared to the real questionnaire filled by actual Social Media user:

Instructions:

This questionnaire consists of 21 groups of statements. Please read each group of statements carefully, and then pick out the one statement in each group that best describes the way you feel. If several statements in the group seem to apply equally well, choose the highest number for that group.

1. Sadness
 0. I do not feel sad.
 1. I feel sad much of the time.
 2. I am sad all the time.
 3. I am so sad or unhappy that I can't stand it.
2. Pessimism
 0. I am not discouraged about my future.
 1. I feel more discouraged about my future than I used to be.
 2. I do not expect things to work out for me.
 3. I feel my future is hopeless and will only get worse.
3. Past Failure
 0. I do not feel like a failure.
 1. I have failed more than I should have.
 2. As I look back, I see a lot of failures.
 3. I feel I am a total failure as a person.
4. Loss of Pleasure
 0. I get as much pleasure as I ever did from the things I enjoy.
 1. I don't enjoy things as much as I used to.
 2. I get very little pleasure from the things I used to enjoy.
 3. I can't get any pleasure from the things I used to enjoy.
5. Guilty Feelings
 0. I don't feel particularly guilty.
 1. I feel guilty over many things I have done or should have done.
 2. I feel quite guilty most of the time.
 3. I feel guilty all of the time.
6. Punishment Feelings
 0. I don't feel I am being punished.
 1. I feel I may be punished.
 2. I expect to be punished.
 3. I feel I am being punished.
7. Self-Dislike
 0. I feel the same about myself as ever.
 1. I have lost confidence in myself.
 2. I am disappointed in myself.
 3. I dislike myself.
8. Self-Criticalness
 0. I don't criticize or blame myself more than usual.
 1. I am more critical of myself than I used to be.
 2. I criticize myself for all of my faults.
 3. I blame myself for everything bad that happens.
9. Suicidal Thoughts or Wishes
 0. I don't have any thoughts of killing myself.
 1. I have thoughts of killing myself, but I would not carry them out.
 2. I would like to kill myself.
 3. I would kill myself if I had the chance.
10. Crying
 0. I don't cry anymore than I used to.
 1. I cry more than I used to.
 2. I cry over every little thing.
 3. I feel like crying, but I can't.
11. Agitation
 0. I am no more restless or wound up than usual.
 1. I feel more restless or wound up than usual.
 2. I am so restless or agitated that it's hard to stay still.
 3. I am so restless or agitated that I have to keep moving or doing something.
12. Loss of Interest
 0. I have not lost interest in other people or activities.
 1. I am less interested in other people or things than before.
 2. I have lost most of my interest in other people or things.
 3. It's hard to get interested in anything.
13. Indecisiveness
 0. I make decisions about as well as ever.
 1. I find it more difficult to make decisions than usual.
 2. I have much greater difficulty in making decisions than I used to.
 3. I have trouble making any decisions.
14. Worthlessness
 0. I do not feel I am worthless.
 1. I don't consider myself as worthwhile and useful as I used to.
 2. I feel more worthless as compared to other people.
 3. I feel utterly worthless.
15. Loss of Energy
 0. I have as much energy as ever.
 1. I have less energy than I used to have.
 2. I don't have enough energy to do very much.
 3. I don't have enough energy to do anything.

Figure 1. Beck's Depression Inventory (part 1)

16. Changes in Sleeping Pattern
0. I have not experienced any change in my sleeping pattern.
1a. I sleep somewhat more than usual.
1b. I sleep somewhat less than usual.
2a. I sleep a lot more than usual.
2b. I sleep a lot less than usual.
3a. I sleep most of the day.
3b. I wake up 1-2 hours early and can't get back to sleep.
17. Irritability
0. I am no more irritable than usual.
1. I am more irritable than usual.
2. I am much more irritable than usual.
3. I am irritable all the time.
18. Changes in Appetite
0. I have not experienced any change in my appetite.
1a. My appetite is somewhat less than usual.
1b. My appetite is somewhat greater than usual.
2a. My appetite is much less than before.
2b. My appetite is much greater than usual.
3a. I have no appetite at all.
3b. I crave food all the time.
19. Concentration Difficulty
0. I can concentrate as well as ever.
1. I can't concentrate as well as usual.
2. It's hard to keep my mind on anything for very long.
3. I find I can't concentrate on anything.
20. Tiredness or Fatigue
0. I am no more tired or fatigued than usual.
1. I get more tired or fatigued more easily than usual.
2. I am too tired or fatigued to do a lot of the things I used to do.
3. I am too tired or fatigued to do most of the things I used to do.
21. Loss of Interest in Sex
0. I have not noticed any recent change in my interest in sex.
1. I am less interested in sex than I used to be.
2. I am much less interested in sex now.
3. I have lost interest in sex completely

Figure 2. Beck's Depression Inventory (part 2)

- Hit Rate (HR). This is a stringent measure that computes the ratio of cases where the automatic questionnaire has exactly the same answer as the real questionnaire. For example, an automatic questionnaire with 5 matches gets HR equal to 5/21 (because there are 21 questions in the form).
Average Hit Rate (AHR): HR averaged over all users.
- Closeness Rate (CR). This measure takes into account that the answers of the depression questionnaire represent an ordinal scale. For example, consider the #17 question:

17. Irritability

- 0. I am no more irritable than usual.
- 1. I am more irritable than usual.
- 2. I am much more irritable than usual.
- 3. I am irritable all the time.

Imagine that the real user answered "0". A system S1 whose answer is "3" should be penalised more than a system S2 whose answer is "1".

For each question, CR computes the absolute difference (*ad*) between the real and the automated answer (e.g. *ad*=3 and *ad*=1 for S1 and S2, respectively) and, next, this absolute difference is transformed into an effectiveness score as follows: $CR = (mad - ad)/mad$, where *mad* is the maximum absolute difference, which is equal to the number of possible answers minus one.

Average Closeness Rate (ACR): CR averaged over all users.

NOTE: in the two questions (#16 and #18) that have seven possible answers {0, 1a, 1b, 2a, 2b, 3a and 3b} the pairs (1a, 1b), (2a, 2b), (3a, 3b) are considered equivalent because they reflect the same depression level. As a consequence, the difference from 3b to 0 is equal to 3 (and the difference from 1a to 1b is equal to 0).

- Difference between overall depression levels (*DODL*).
The previous measures assess the systems' ability to answer each question in the form. This measure, instead, does not look at question-level hits or differences but computes the overall depression level (sum of all the answers) for the real and automated questionnaire and, next, the absolute difference (*ad_overall*) between the real and the automated score is computed.
Depression levels are integers between 0 and 63 and, thus, *DODL* is normalised into [0,1] as follows:
 $DODL = (63 - ad_overall)/63$
average DODL (ADODL): *DODL* averaged over all users.
- in the psychological domain, it is customary to associate depression levels with the following categories:

- minimal depression (depression levels 0-9)
- mild depression (depression levels 10-18)
- moderate depression (depression levels 19-29)
- severe depression (depression levels 30-63)

The last effectiveness measure consists of computing the fraction of cases where the automated questionnaire led to a depression category that is equivalent to the depression category obtained from the real questionnaire. This measure will be referred to as **DCHR (DEPRESSION CATEGORY HIT RATE)**.

3 Task 3: results

Run	AHR	ACR	ADODL	DCHR
BioInfo@UAVR	34.05%	66.43%	77.70%	25.00%
BiTeM	32.14%	62.62%	72.62%	25.00%
CAMH_GPT_nearest_unsupervised	23.81%	57.06%	81.03%	45.00%
CAMH_GPT_supervised.181_features.58hr	35.47%	68.33%	75.63%	20.00%
CAMH_GPT_supervised.769_features.55hr	36.43%	67.22%	72.30%	20.00%
CAMH_GPT_supervised.949_features.75hr	36.91%	69.13%	75.63%	15.00%
CAMH_LIWC_supervised_SVM	35.95%	66.59%	75.48%	25.00%
Fazl	22.38%	56.27%	72.78%	5.00%
Illinois	22.62%	56.19%	66.35%	40.00%
ISIKol_multiSimilarity-5000-Dtac-Qtac	29.76%	57.94%	74.13%	25.00%
ISIKol-bm25-1.2-0.75-5000-Dtac-Qtac	29.76%	57.06%	72.78%	25.00%
ISIKol-lm-d-1.0-5000-Dtac-Qtac	30.00%	57.94%	73.02%	15.00%
Kimberly	38.33%	64.44%	66.19%	20.00%
UNSLA	37.38%	67.94%	72.86%	30.00%
UNSLB	36.93%	70.16%	76.83%	30.00%
UNSLC	41.43%	69.13%	78.02%	40.00%
UNSLD	38.10%	67.22%	78.02%	30.00%
UNSLE	40.71%	71.27%	80.48%	35.00%