

Secondo Progetto per il corso di Social Computing A.A. 2022/23

Obiettivi

L'obiettivo generale del progetto consiste nel generare e dispiegare un task di crowdsourcing utilizzando il software Crowd_Frame e nel raccogliere ed analizzare dati. Il task ha il triplice scopo di: (i) raccogliere giudizi di verità relativi a frasi (statement); (ii) stimare la qualità di spiegazioni (explanations) generate da modelli di machine learning; (iii) individuare evidenze al fine di supportare ed eventualmente aggiornare la scelta del giudizio di verità. Il task deve permettere ai worker di fornire ciascun giudizio di verità utilizzando una scala di valutazione categoriale ordinale. I worker forniscono la stima della qualità di ciascuna spiegazione utilizzando una scala intervallare. Inoltre, individuano evidenze utilizzando il motore di ricerca personalizzato. Tali evidenze devono permettere di aggiornare il giudizio di verità, qualora necessario, riproponendo la scala di valutazione utilizzata. Gli statement provengono dal dataset FEVER [1] (<https://fever.ai/dataset/fever.html>), acronimo di Fact Extraction and VERification. Le spiegazioni provengono dal dataset e-FEVER [2], disponibile pubblicamente su richiesta agli autori. La scala di valutazione per i giudizi di verità è simile a quella utilizzata da PolitiFact [3], noto sito di fact-checking.

Cosa fare

1. Inviare una mail per gruppo con oggetto: [Progetto SocCom 2 - 2022/2023 - Dataset] a michael.soprano@uniud.it per ricevere il dataset composto da un file csv di 3 righe.
Nota: non divulgare pubblicamente il dataset ricevuto, poiché è necessario il permesso degli autori.
2. Ogni riga descrive uno statement relativo ad un argomento, riportando 5 informazioni:
 - a. id: identificatore dello statement
 - b. statement: testo dello statement
 - c. explanation_human: spiegazione dello statement curata da esperti
 - d. explanation_model: spiegazione dello statement generata da un modello di machine learning
 - e. label: etichetta che indica se la spiegazione supporta (SUPPORTS) o confuta (REFUTES) lo statement

La tabella seguente mostra una riga d'esempio del dataset fornito.

id	statement	explanation_hu man	explanation_mo del	label
94151	Uranium-235 was discovered by at least one person.	It was discovered in 1935 by Arthur Jeffrey Dempster. Arthur Jeffrey Dempster (August 14, 1886 - March 11, 1950) was a	Uranium-235 was discovered by Arthur Jeffrey Dempster in 1935. The claim is true.	SUPPORTS

		Canadian-American physicist best known for his work in mass spectrometry and his discovery of the uranium isotope 235U .		
--	--	--	--	--

3. Scrivete un Jupyter Notebook al fine di creare gli HITs del vostro task a partire dal dataset ricevuto, con le seguenti caratteristiche:

- Ogni HIT contiene 3 elementi
- Ogni elemento di un HIT ha 4 attributi:
 - id: identificatore dello statement
 - statement: testo dello statement
 - explanation: testo della spiegazione
 - label: etichetta della spiegazione
- Due elementi su tre (2/3) di un HIT assegnano all'attributo explanation il valore di explanation_model
- Un elemento su tre (1/3) di un HIT assegnano all'attributo explanation il valore di explanation_human
- Ogni valore di explanation_model dev'essere valutato da 3 worker diversi
- Ogni valore di explanation_human dev'essere valutato da 2 worker diversi
- La posizione dei 3 elementi in un HIT deve essere casuale
- Il numero totale di HIT è pari a 12

Si supponga che E1, E2, E3 indichino le spiegazioni generate, mentre G1, G2 e G3 quelle fornite da esseri umani. La seguente tabella mostra un possibile schema valido di allocazione di sei dei dodici HITs richiesti:

Posizione 1	Posizione 2	Posizione 3
E1	G1	E2
G1	E3	E1
E2	E1	G2
E2	G2	E3
E3	E1	G3
G3	E2	E3
...

4. Progettate e dispiegate il task di crowdsourcing con le seguenti caratteristiche:

- Visualizzate i seguenti attributi per ciascun elemento:
 - statement
 - label
 - explanation
- Configurate l'annotazione di testo con una sola etichetta:

- *Testo*: Select
- *Colore*: un qualsiasi colore HTML, es: #5f9ea0
- c. Abilitate l'annotazione del seguente attributo e rendetela obbligatoria:
 - *explanation*
- d. Aggiungete una prima dimensione di valutazione con le seguenti caratteristiche:
 - *Scopo*: indicare il valore di verità dello statement
 - *Nome*: truthfulness_1
 - *Scala di valutazione*: categoriale ordinale a 6 valori.
 - Lie (0)
 - False (1)
 - Barely True (2)
 - Half True (3)
 - Mostly True (4)
 - True (5)
 - *Stile*:
 - Tipo: list
 - Posizione: top
- e. Aggiungete una seconda dimensione di valutazione con le seguenti caratteristiche:
 - *Scopo*: stimare la qualità della spiegazione nel supportare la scelta del valore di verità
 - *Nome*: explanation_quality
 - *Scala di valutazione*: intervallare
 - Minimo: 1
 - Massimo: 100
 - Step: 1
 - *Stile*:
 - Tipo: list
 - Posizione: middle
- f. Aggiungete una terza dimensione di valutazione con le seguenti caratteristiche:
 - *Scopo*:
 - Cercare evidenza a supporto del giudizio di verità fornito
 - Aggiornare il giudizio di verità sulla base dell'eventuale evidenza trovata
 - *Nome*: truthfulness_2
 - *Scala di valutazione*: categoriale ordinale a 6 valori
 - Lie (0)
 - False (1)
 - Barely True (2)
 - Half True (3)
 - Mostly True (4)
 - True (5)
 - *Richiesta di URL*
 - Provider: BingWebSearch
 - *Richiesta di giustificazione*

- Parole minime: 5
 - *Stile*:
 - Tipo: list
 - Posizione: bottom
- g. Impostazioni:
 - Ogni worker deve passare almeno X secondi su ciascun elemento (scegliete voi X)
 - Ogni worker ha Y tentativi a disposizione per svolgere il task (scegliete voi Y)
- h. Istruzioni:
 - Scrivete istruzioni generali adeguate
 - Scrivete istruzioni di valutazione adeguate
- 5. Dispiegate il task utilizzando lo script di inizializzazione descritto a lezione
- 6. Dispiegate il task attraverso la Requester Sandbox di Amazon Mechanical Turk secondo le procedure descritte a lezione
- 7. Fate svolgere tutti e 12 gli HIT del vostro task chiedendo ai membri di 4 gruppi di lavorare per voi come worker. Ciascun worker deve accedere al task attraverso la Worker Sandbox di Amazon Mechanical Turk
- 8. Scaricate i dati prodotti dai worker utilizzando lo script di download descritto a lezione
- 9. Calcolate il *Percent Agreement* tra worker per i valori forniti rispetto a ciascuna dimensione categoriale
- 10. Calcolate la percentuale media di testo annotato per ciascuna spiegazione
 - a. Ordinate le spiegazioni sulla base di tale parametro
- 11. Determinate il numero di volte in cui i worker aggiornano l'annotazione proposta per ciascuna spiegazione
 - a. Costruite un ranking delle spiegazioni sulla base di tale parametro
- 12. Calcolate il tempo medio impiegato dai worker per valutare ciascun elemento
- 13. Aggiungete eventuali altre analisi che ritenete opportune ed utili
- 14. Descrivete brevemente le considerazioni tratte dall'analisi dei dati prodotti dai worker

Informazioni aggiuntive

- Generate una chiave per le API di Bing utilizzando il piano gratuito ed impostatela come valore della variabile d'ambiente `bing_api_key`
- Aggiungete altre istruzioni oltre a quelle generali e di valutazione dove necessario, prendendo spunto dalla configurazione del task d'esempio visto a lezione
- Esplorate le configurazioni d'esempio fornite a lezione per capire quale parte di esse porta all'inizializzazione di ciascun elemento dell'interfaccia utente

Come consegnare

1. I gruppi **devono** essere formati da **tre** persone (i gruppi più o meno numerosi verranno penalizzati)
2. Si devono consegnare i seguenti elementi:
 - Relazione di al massimo 6 pagine
 - Scrivete anche i vostri nomi cognomi e numeri di matricola
 - Dataset di input assegnato
 - Rinominate il file aggiungendo i vostri cognomi come suffisso nel formato: `-cognome1_cognome2_cognome3`
 - Esempio: `group_1-cognome1_cognome2_cognome3.csv`
 - Configurazione del task dispiegato

- Scrivete l'URL del task spiegato nella relazione
 - Shapshot, Dataframe, etc. prodotti scaricando i risultati dopo lo svolgimento del task
 - Risultati prodotti durante la fase di analisi dei dati
- 3. Consegnate via mail a tutti i docenti (un unico messaggio)
 - mizzaro@uniud.it
 - michael.soprano@uniud.it
 - david.labarbera@uniud.it
 - Oggetto della mail nel formato:
[Progetto SocCom 2 - 2022/2023 - Consegna]
cognome1_cognome2_cognome3
 - in allegato alla mail un unico file zippato che quando scompattato produce una singola cartella con nome cognome1_cognome2_cognome3
- 4. Punteggio:
 - 5 punti in trentesimi per i migliori 25%,
 - 4 punti per i seguenti 25%,
 - 3 punti per i seguenti 20%,
 - 2 punti seguenti 15%,
 - 1 punto per i seguenti 15%,
 - 0 punti a discrezione dei docenti per progetti non adeguati o per chi non consegna
- 5. **Scadenza: Sabato 14 Gennaio 2023 AoE Timezone**

Bibliografia

- [1] James Thorne, Andreas Vlachos, Christos Christodoulopoulos, and Arpit Mittal. 2018. FEVER: a Large-scale Dataset for Fact Extraction and VERification. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 809–819, New Orleans, Louisiana. Association for Computational Linguistics. <http://dx.doi.org/10.18653/v1/N18-1074>
- [2] Dominik Stammach, Elliot Ash. 2020. e-FEVER: Explanations and Summaries for Automated Fact Checking. *Conference for Truth and Trust Online (TTO 2020)*, Online, October 16-17. <https://doi.org/10.3929/ethz-b-000453826>
- [3] AA. VV. The Principles of the Truth-O-Meter: PolitiFact's methodology for independent fact-checking. <https://www.politifact.com/article/2018/feb/12/principles-truth-o-meter-politifact-methodology-i/>