

Patient Knowledge Extraction and Representation from Unlabeled Social Posts

Giacomo Frisoni*, Gianluca Moro, Antonella Carbonaro

Department of Computer Science and Engineering – DISI, University of Bologna, Italy



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Introduction

Rare diseases pose particular challenges to patients, families, caregivers, clinicians and researchers. Due to the scarce availability of information and their disintegration, in recent years we are witnessing a strong growth of patient communities on social platforms such as Facebook. Within them, users share experiences and seek answers to questions aimed at improving their health conditions (e.g., “**which are the most effective medical treatments from patients’ viewpoint?**”). Though the unlabeled data generated in this context are of high value, the currently existing ontologies and resources tend to ignore them. Moreover, the explanation of a phenomenon from text documents is still a difficult and poorly addressed problem.

Aims

- A novel methodology of descriptive text mining, capable of offering accurate explanations in unsupervised settings, of discovering relationships among concepts and quantifying them according to their statistical significance (e.g., “citrus fruit” \leftrightarrow “acid reflux”: 87%).
- Introduction of a new knowledge graph learning approach.
- Effectiveness demonstration through a case study focused on Esophageal Achalasia (ORPHA: 930), extending ORDO and introducing a patient-centered knowledge graph into the world of Linked Open Data.

Patient Knowledge Extraction

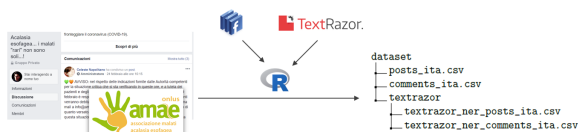


Figure 1: From Social Data to Dataset construction

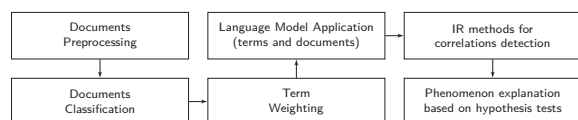


Figure 2: Main modules of the methodology

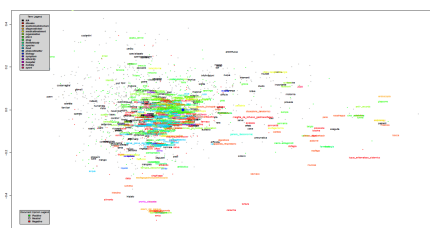


Figure 3: Modeling of patient posts and terms in a latent semantic space

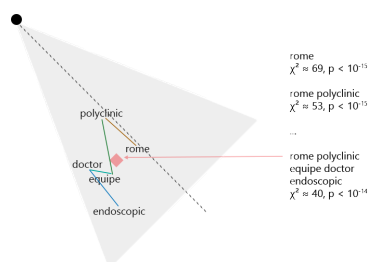


Figure 4: Iterative approach for phenomenon explanation

Patient Knowledge Representation

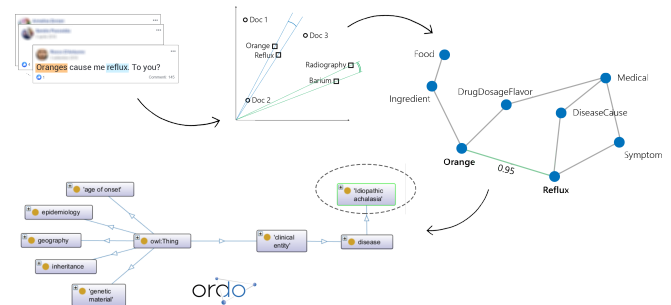


Figure 5: Patient Knowledge Graph Learning example for Esophageal Achalasia

Experiments and Results

- **Explanations effectiveness.** Successful recognition of the positive and negative aspects of the two main intervention techniques for Esophageal Achalasia (Heller-Dor and POEM) through local analyzes and p-value < 0.01 (Table 1).

	Pos Explanation	Neg Explanation
Heller-Dor	equipe, dr, mario, costantini, salvador, padua, antireflux, plastic	problems, drink, eat
POEM	rome, prof, gemelli, costamagna, equipe, familiari, lombardy	reflux, problems, liquid, pain, inflammation, antacids

Table 1: Explanations for positive and negative opinions about Achalasia treatments

- **Correlations evaluation.** $\approx 78\%$ accuracy in identifying gold standard correlations (e.g., symptom \leftrightarrow drug, place \leftrightarrow doctor), both positive and negative, known in the literature and reported by domain experts (Table 2).

P-value Threshold	Confusion Matrix				Statistical Indices			
	TP	TN	FP	FN	ACC	PRE	TPR	FPR
0.7	101	68	36	19	75.45	73.72	84.17	34.62
0.8	96	78	26	24	77.68	78.69	80.00	21.31
0.9	72	98	6	48	75.89	92.31	60.00	5.77

Table 2: Confusion matrices and relative statistical indices obtained with three different acceptance thresholds. ACC=Accuracy, PRE=Precision, TPR=TruePositiveRate, FPR=FalsePositiveRate

- Discovery of new correlations, currently being examined by experts, such as PPI \rightarrow negative opinion.

Conclusions and Future Works

- Combined use of Text Mining and Semantic Web techniques, with a modular, automatic and domain-independent KE e KR methodology.
- Identification of statistical evidences and scientific medical correlations directly from patients’ posts accumulated over time, avoiding their manual reading.
- The methodology could be applied to different communities of patients for Esophageal Achalasia (comparing the results) and to other diseases.

References

This poster abstract is a reworking of two contributions.

Frisoni, G., Moro, G., & Carbonaro, A. (2020a). Discovering Significant Medical Correlations about Diseases from Unlabeled Social Posts: A Novel Methodology of Descriptive Text Mining. *Submitted to DATA*.

Frisoni, G., Moro, G., & Carbonaro, A. (2020b). Towards Rare Disease Knowledge Graph Learning from Social Posts of Patients. *Submitted to Rii Forum*.