

AI for Health and Medicine

Intelligent Information Mining - Research Group

S. Faralli, L. Madeddu, G. Stilo, P. Velardi
Intelligent Information Mining - CINI Roma
<http://iim.di.uniroma1.it> - velardi@di.uniroma1.it

Abstract

Il gruppo di ricerca *Intelligent Information Mining* nodo CINI di Roma, tra i vari interessi, svolge attività di ricerca nel settore *AI for Health and Medicine*. I membri del gruppo interessati a questa tematica sono: S. Faralli (ricercatore), L. Madeddu (studente di dottorato), G. Stilo (ricercatore) e P. Velardi (professore ordinario),

1 Collaborazioni

Il gruppo di ricerca ha in essere le seguenti collaborazioni:

- Partecipa al collegio di **Dottorato di Ricerca in Tecnologie Biomediche Innovative in Medicina Clinica**¹ del Dipartimento di Medicina traslazionale e di precisione - Sapienza.
- Collabora con la **Fondazione Francesco Balsano**², che promuove la ricerca scientifica in particolare nel campo delle indagini cliniche e sperimentali ed ha esclusivamente carattere di pubblica utilità³;
- Coordinatore del gruppo di lavoro *Network Medicine Working Group: Methods* promosso da **Sapienza information-based Technology Innovation Center for Health**³ e **Harvard Medical School**.

2 Attività editoriali e organizzative

Il gruppo di lavoro è coinvolto nell'organizzazione del primo workshop internazionale *Virtual Reality and Artificial Intelligence in Medical Applications (VRAIMA 2019)*⁴ che si terrà il 18 giugno 2019 a Valencia, Spagna, e nella *Interdisciplinary Research Track on Artificial Intelligence and Medicine* della 6th ACM *womenEncourage*, che si terrà a Roma dal 16 al 18 Settembre 2019⁵.

¹<https://www.uniroma1.it/offerta-formativa/dottorato/2019/tecnologie-biomediche-innovative-medicina-clinica>

²<http://www.fondazionefrancescobalsano.it>

³<https://web.uniroma1.it/stitch/>

⁴<https://sites.google.com/unitelmasapienza.it/vraima2019>

⁵<https://womencourage.acm.org/2019/>

3 Progetti di Ricerca

Deep Graph-based Method for Disease Gene Prediction:

Nell'ultima decade, il campo della ricerca biomedica si è focalizzato sull'analisi delle complesse relazioni fra geni e malattie [Menche *et al.*, 2015] [Agrawal *et al.*, 2018] hanno mostrato come approcci predittivi [Ghiassian *et al.*, 2015; Silberberg *et al.*, 2017] basati su una rete di proteine, siano efficaci per la scoperta di relazioni, ancora sconosciute, fra geni e malattie. Tuttavia, l'incompletezza della rete (l'interattoma umano) è il limite principale alle capacità predittive di tali metodi. Attualmente si stima che solo il 20% dell'Interattoma umano sia noto [Venkatesan *et al.*, 2009]. Attualmente il gruppo sta lavorando ad un approccio basato su metodi deep learning capace di sfruttare le caratteristiche latenti presenti nell'Interattoma umano, per il riconoscimento di nuovi disease-genes. Diversamente da altri studi basati sulla sola connessione tra nodi nell'Interattoma umano, il nostro modello cattura dalla rete informazioni che caratterizzano le regolarità strutturali, l'interconnessione delle proteine, le regolarità funzionali, e le proprietà biologiche (tessuti, malattie, pathways etc.). I risultati ottenuti mostrano che la combinazione delle informazioni biologiche associate alle proteine con la topologia della rete, permette un significativo miglioramento delle capacità predittive di tali sistemi.

The Social Phenotype of Diseases: Methods and Application to Diabetic Patients:

Vivere nell'era digitale ha alterato la nostra visione dell'assistenza sanitaria. Negli ultimi anni è emerso un crescente interesse medico focalizzato sul paziente e la sua esperienza clinica. Al contempo, attraverso Internet, le persone generano un'immensa quantità di dati riguardanti la propria salute [Greaves *et al.*, 2013]. Ne sono esempio siti web e forum in cui i pazienti sono a diretto contatto con figure mediche per esprimere dubbi, problemi e domande sulle loro esperienze. In tale circostanza, l'analisi dei Big Data clinici rappresenta un'importante opportunità per perfezionare la tempestività e l'efficienza del sistema sanitario [Kavakiotis *et al.*, 2017]. Lo studio effettuato analizza lo stato di salute di utenti affetti da diabete mellito, definendone i "digital phenotypes" [Jain *et al.*, 2015]. Si vuole quindi caratterizzare, attraverso l'analisi di siti di salute, l'esperienza diretta di pazienti diabetici, in modo da fornire un efficace supporto per il settore clinico e farmaceutico. La ricerca, inizialmente si concentra sulla descrizione analitica e statistica

dei dati raccolti e successivamente si focalizza nell'estrazione e la classificazione, tramite algoritmi di Clustering e Machine Learning, delle principali tematiche emerse.

Improving Quality of Life for Older Adults: Il gruppo di ricerca sviluppa sistemi di *live speech emotion detection* per la piattaforma di storytelling per adulti e anziani *VoicingElder* [Ryu, 2017].⁶ *VoicingElder* utilizza avatar interattivi per riflettere i dialoghi e i gesti dal vivo. Gli utenti recitano rappresentazioni di se stessi, o di personaggi di fantasia in quattro fasi di sviluppo (bambino, adolescente, adulto e anziano), modellando e riflettendo la loro storia di vita in modo giocoso e coinvolgente. *VoicingElder* promuove la ricerca transdisciplinare in gerontologia, salute, arte, psicologia e tecnologia. In particolare in questo contesto vengono sperimentate e sviluppate tecniche di *emotion detection* mediante metodologie di *aspect-based sentiment analysis* su flussi di frammenti di testo che vengono generati in "tempo reale" da motori di *speech to text*.

Twitter mining for fine-grained syndromic surveillance: I messaggi digitali prodotti dagli utenti internet possono rappresentare un enorme set di dati utili a fornire informazioni (in tempo reale) utili ai sistemi di sorveglianza sindromica. Il linguaggio utilizzato dalle persone differisce da quello medico, quindi riuscire ad interpretare la terminologia dei pazienti risulta un aspetto cruciale per estrarre informazioni inerenti alla salute. In questo filone di ricerca possiamo annoverare:

- In (Twitter mining for fine-grained syndromic surveillance) [Velardi *et al.*, 2014] viene presentata una metodologia per la diagnosi precoce e l'analisi delle epidemie basate sui messaggi di Twitter.
- In (Can Twitter be a source of information on allergy? Correlation of pollen counts with tweets reporting symptoms of allergic rhinoconjunctivitis and names of antihistamine drugs) [Gesualdo *et al.*, 2015] i dati derivati da Twitter che identificano sindromi di rinocongiuntivite congiunti all'utilizzo di un farmaco antistaminico sono utilizzati come segnale da comparare le tendenze settimanali del numero di pollini raccolti dal National Allergy Bureau.
- In (Predicting flu epidemics using Twitter and historical data) [Stilo *et al.*, 2014] le serie temporali dei messaggi Twitter (che riportano i sintomi dell'influenza-like (ILI)) vengono utilizzati per la previsione delle stagioni influenzali attuali e passate (17 stagioni complessive) raccolte dal CDC dell' U.S. Department of Health & Human Services.

Riferimenti bibliografici

- [Agrawal *et al.*, 2018] Monica Agrawal, Marinka Zitnik, e Jure Leskovec. Large-scale analysis of disease pathways in the human interactome. In *Pacific Symposium on Biocomputing*, volume 23, pages 111–122, 2018.
- [Gesualdo *et al.*, 2015] Francesco Gesualdo, Giovanni Stilo, Angelo D'Ambrosio, Emanuela Carloni, Elisabetta Pandolfi, Paola Velardi, Alessandro Fiocchi, e Alberto E. Tozzi. Can twitter be a source of information on allergy? correlation of pollen counts with tweets reporting symptoms

of allergic rhinoconjunctivitis and names of antihistamine drugs. *PLOS ONE*, 10(7):1–11, 07 2015.

- [Ghiassian *et al.*, 2015] Susan Dina Ghiassian, Jörg Menche, e Albert-László Barabási. A disease module detection (diamond) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Computational Biology*, 11(4), 2015.
- [Greaves *et al.*, 2013] Felix Greaves, Daniel Ramirez-Cano, Christopher Millett, Ara Darzi, e Liam Donaldson. Harnessing the cloud of patient experience: Using social media to detect poor quality healthcare. *BMJ quality & safety*, 22, 01 2013.
- [Jain *et al.*, 2015] Sachin Jain, Brian Powers, Jared B Hawkins, e John S Brownstein. The digital phenotype. *Nature biotechnology*, 33:462–463, 05 2015.
- [Kavakiotis *et al.*, 2017] Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, N Maglaveras, I Vlahavas, e Ioanna Chouvarda. Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 01 2017.
- [Menche *et al.*, 2015] Jörg Menche, Amitabh Sharma, Maksim Kitsak, Susan Dina Ghiassian, Marc Vidal, Joseph Loscalzo, e Albert-László Barabási. Uncovering disease-disease relationships through the incomplete interactome. *Science*, 347(6224), 2015.
- [Ryu, 2017] Semi Ryu. Avatar life-review: Seniors reminiscing through virtual bodies. In Constantine Stephanidis, editor, *HCI International 2017 – Posters' Extended Abstracts*, pages 488–496, Cham, 2017. Springer International Publishing.
- [Silberberg *et al.*, 2017] Yael Silberberg, Martin Kupiec, e Roded Sharan. Gladiator: a global approach for elucidating disease modules. *Genome Medicine*, 9, 05 2017.
- [Stilo *et al.*, 2014] Giovanni Stilo, Paola Velardi, Alberto E. Tozzi, e Francesco Gesualdo. Predicting flu epidemics using twitter and historical data. In Dominik Ślezak, Ah-Hwee Tan, James F. Peters, e Lars Schwabe, editors, *Brain Informatics and Health*, pages 164–177, Cham, 2014. Springer International Publishing.
- [Velardi *et al.*, 2014] Paola Velardi, Giovanni Stilo, Alberto E. Tozzi, e Francesco Gesualdo. Twitter mining for fine-grained syndromic surveillance. *Artificial intelligence in medicine*, 61 3:153–63, 2014.
- [Venkatesan *et al.*, 2009] Kavitha Venkatesan, Jean-Francois Rual, Alexei Vazquez, Ulrich Stelzl, Irma Lemmens, Tomoko Hirozane-Kishikawa, Tong Hao, Martina Zenkner, Xiaofeng Xin, Kwang-II Goh, Muhammed A Yildirim, Nicolas Simonis, Kathrin Heinzmann, Fana Gebreab, Julie M Sahalie, Sebiha Cevik, Christophe Simon, Anne-Sophie de Smet, Elizabeth Dann, e Marc Vidal. An empirical framework for binary interactome mapping. *Nature methods*, 6:83–90, 01 2009.

⁶<http://voicingelder.com/>