Valerio La Gatta

eXplainable Artificial Intelligence for multimodal social media analysis

Tutor: Prof. Vincenzo Moscato

Cycle: XXXVI   Year: Second

My background

- MSc degree in Computer Engineering from University of Naples Federico II

- Research group: Pattern Analysis and Intelligence Computation for mUltimedia System (PICUS)

- PhD start date: 01/11/2020

- Scholarship type: Unina

Research field of interest

- Multimodal fact-checking on social media platforms
  - Support manual fact-checkers work and automate the verification process
  - The role of multimodal information (e.g. text, image, audio) within the fact-checking pipeline
  - The spread of false claims during major geopolitical events

- Multimodal harmful content classification
  - Hateful Internet meme detection
  - Leverage external knowledge to improve the detection process

Summary of study activities

- Ad hoc PhD courses:

  – Big Data Architecture and Analytics, Prof. Giancarlo Sperlì

- Courses borrowed from MSc curricula:

  – Web and Real Time Communication Systems, Prof. Simon Pietro Romano

- Research:

  – Started an internship at University of Southern California (USC), Los Angeles, under the supervision of Prof. Emilio Ferrara

Research activity: fact-checking

- *Context*: the fact-checking pipeline
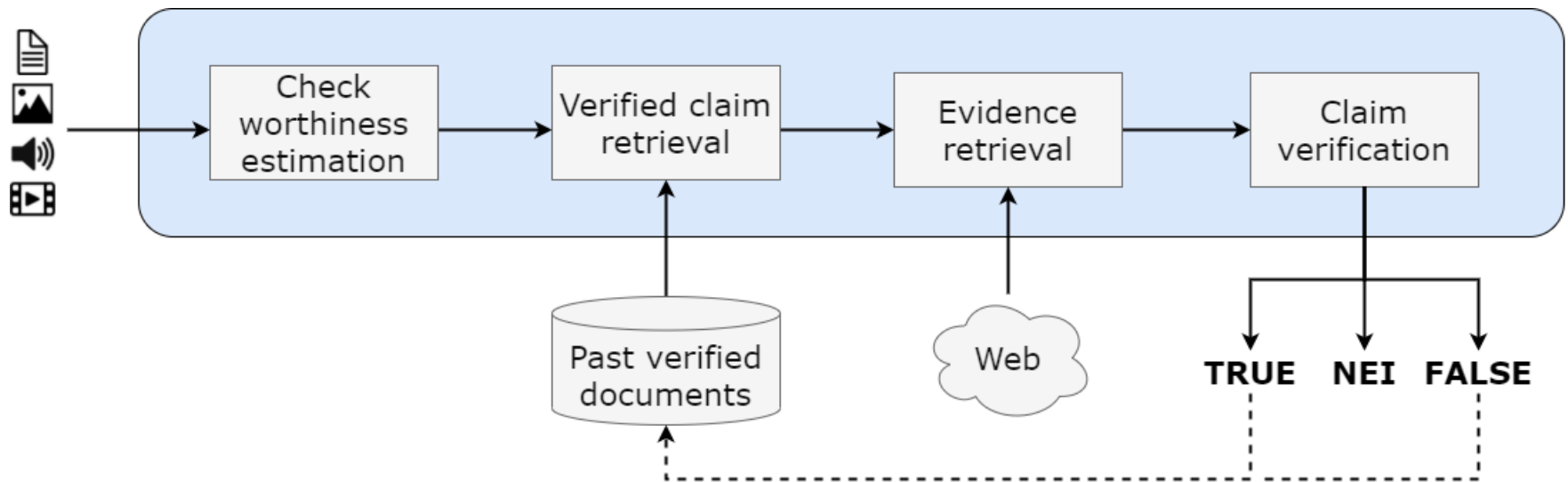
Research activity: fact-checking

- *Problem*: verified claim retrieval, a.k.a. detecting previously fact-checked information
  - Increase manual fact-checkers productivity
  - Improve automatic fact-checking
- *Objective*: retrieval and re-ranking a list of verified documents according to the relevance with an input claim.
  - Multimodal settings include the text and the image(s) of both the input claim and the verified documents

Research activity: fact-checking

- *Methodology:* a two-stages retriever-reranker pipeline
  - Efficient retriever scalable to million-scale (verified) document corpus.
  - Powerful re-ranker to order the retrieved documents.
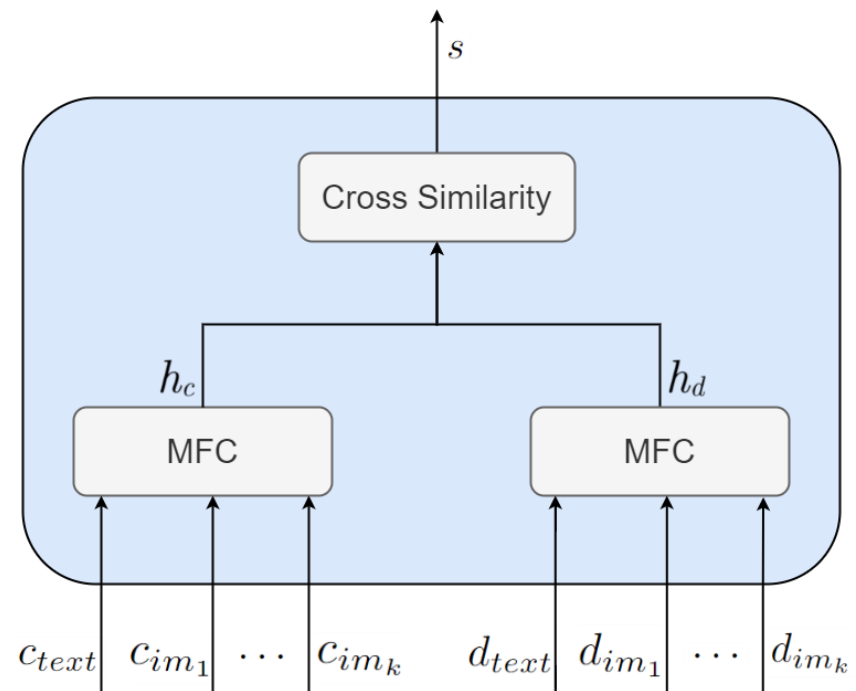
Research activity: fact-checking

- *Results:*

  – Outperform all state-of-the-art re-ranking models

  – Only competitor which (slighlty) overcomes the baseline

Method	MM	Politifact				
		HIT@3	HIT@5	NDCG@1	NDCG@3	NDCG@5
BM25		.379	.433	.182	.292	.313
MatchPyramid		.455	.503	.294	.389	.408
KNRM		.636	.722	.422	.549	.585
BERT		.786	.856	.505	.675	.704
MAN	✓	.732	.786	.551	.654	.676
NSMN		.551	.679	.379	.477	.531
sentence-BERT		.139	.176	.059	.098	.113
Ours	✓	.918	.922	.701	.712	.721

Politifact
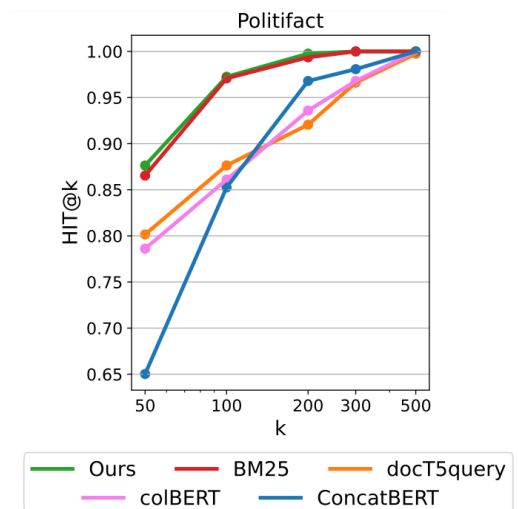
Research activity: meme detection

- *Context:* hateful Internet meme detection
- *Problem:*
  - Text and image modalities within a meme are not always semantically consistent.
 - The understanding of a meme often relies on our background knowledge.

Research activity: meme detection

- *Objective:*
  - Retrieve and match external background knowledge related to both the textual and visual modalities
 - Inject such knowledge into the detection task

Research activity: meme detection

- *Methodology:*
  - Build a knowledge graph using the visual and textual entities extracted from the meme and a common-sense knowledge base (e.g. ConceptNet)
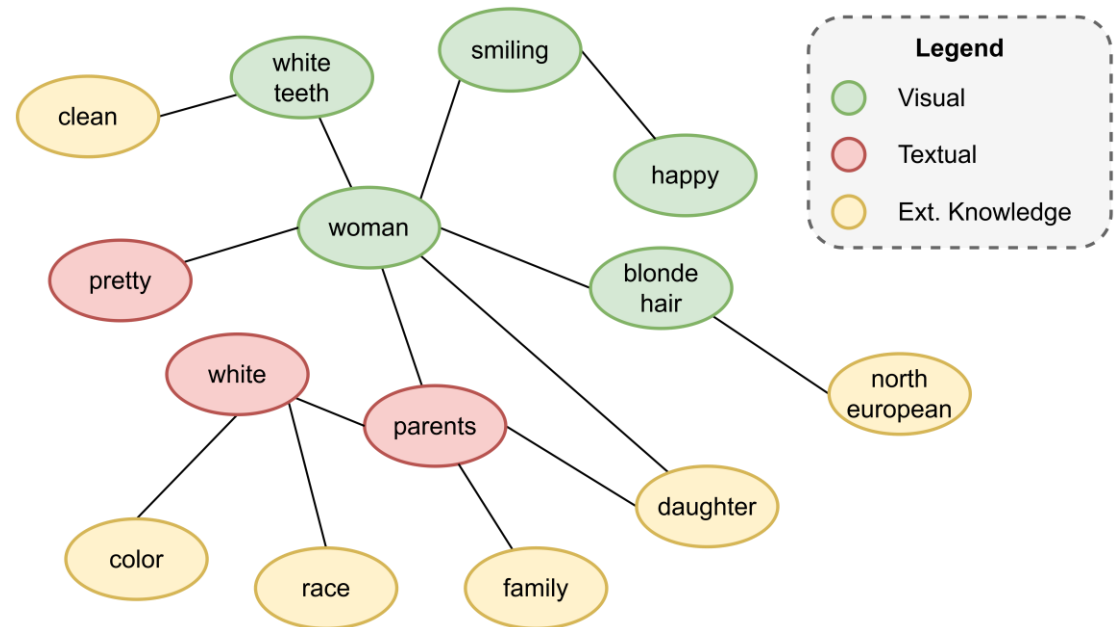
Research activity: meme detection

- *Methodology:*
  - Design a reasoning step which dinamycally learns which (subset of) knowledge is most useful to detect the hatefulness of the meme
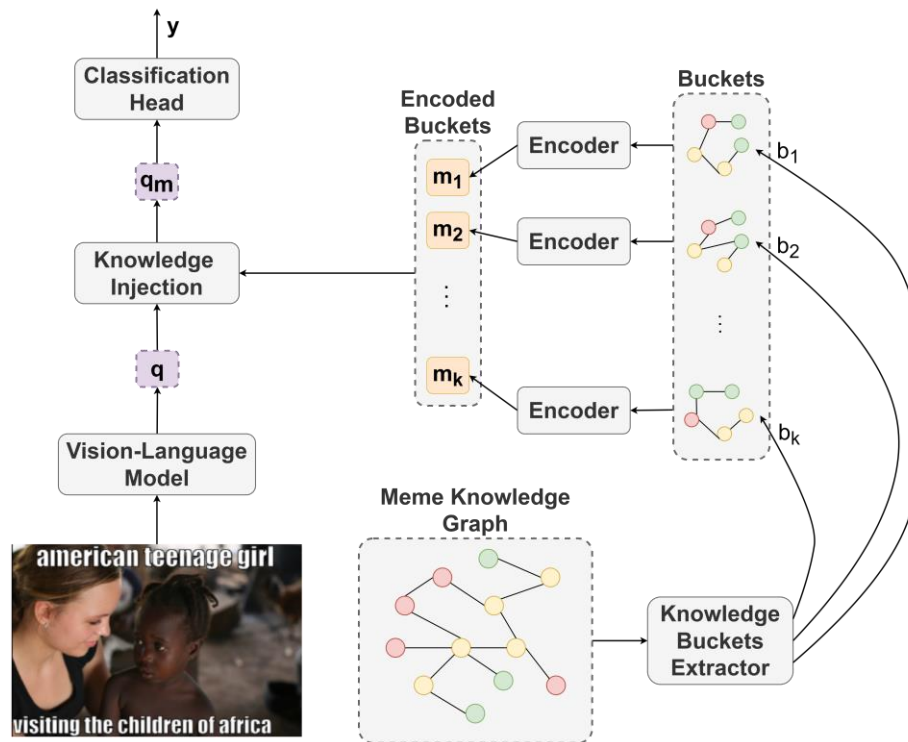
Research activity: meme detection

- (Preliminary) results:
  - The baseline performs image and text embedding with ResNet and BERT, respectively, followed by a classification head.

Model	Facebook Hateful Meme (AUC)	Misogyny Detection (F1-score)
Baseline	0.721	0.815
Baseline + Captioning	0.701	0.818
Baseline + Graph embedding	0.732	0.822
Baseline + Graph reasoning	**0.739**	**0.834**

Products

[P1]	**V. La Gatta**, V. Moscato, M. Pennone, M. Postiglione, G. Sperlì; "Music Recommendation via Hypergraph Embedding"; IEEE Transactions on Neural Networks and Learning Systems, IEEE TNNLS; published
[P2]	A. Ferraro, A. Galli, **V. La Gatta**, V. Moscato, M. Postiglione, G. Sperlì; "An epidemiological Neural Network model exploiting dynamic graph structured data"; IEEE World Congress on Computational Intelligence (IEEE WCCI2022); published
[P3]	A. Barducci, S. Iannaccone, **V. La Gatta**, V. Moscato, G. Sperlì, S. Zavota; "An end-to-end framework for information extraction from Italian resumes"; Expert Systems with Applications, ESWA; published
[P4]	A. Ferraro, A. Galli, **V. La Gatta**, M. Postiglione; "A Deep Learning pipeline for Network Anomaly Detection based on Autoencoders"; IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence, and Neural Engineering; IEEE MetroXRAINE 2022; published

Products

[P5]	T. Chakraborty, **V. La Gatta**, V. Moscato, G. Sperli; "Information retrieval algorithms and neural ranking models to detect previously fact-checked information"; Neurocomputing; submitted
[P6]	R. Formisano, **V. La Gatta**, V. Moscato, G. Sperli; "A Novel Multimodal Retrieval System for Previously Fact-checked Information Detection"; ACM Transactions on Management Information Systems, ACM TMIS; submitted
[P7]	A. Galli, **V. La Gatta**, V. Moscato, M. Postiglione, G. Sperlì; "Interpretability in AI-based Behavioral Malware Detection Systems", IEEE Transactions on Dependable and Secure Computing, IEEE TDSC; submitted

Tutorship

- Co-supervisor of seven master theses in Computer Engineering regarding automatic fact-checking, hateful meme detection and multimodal fake news detection

- Weekly two hours of teaching activities regarding practical lectures/seminars during the courses of "Big Data Engineering" (Master Degree in Computer Engineering) and "Machine Learning and Big Data per la salute" (Master Degree in Biomedical Engineering)

Next year

- Extend the knowledge injection approach to other multimodal disinformation classification tasks (e.g. fake news detection, stance detection)

- Leverage external knowledge to explain the (hateful) classification

- Study on cross-platform harmful content diffusion (in collaboration with USC)

- Draft topic of the thesis:
  - Multimodal disinformation mining

Thank you for the attention!