Automatic analysis of Rapid Astronomical Bulletins using knowledge graphs and natural language processing

Andrei Variu
Andrii Neronov, Volodymyr Savchenko

AstroORDAS

MMODA

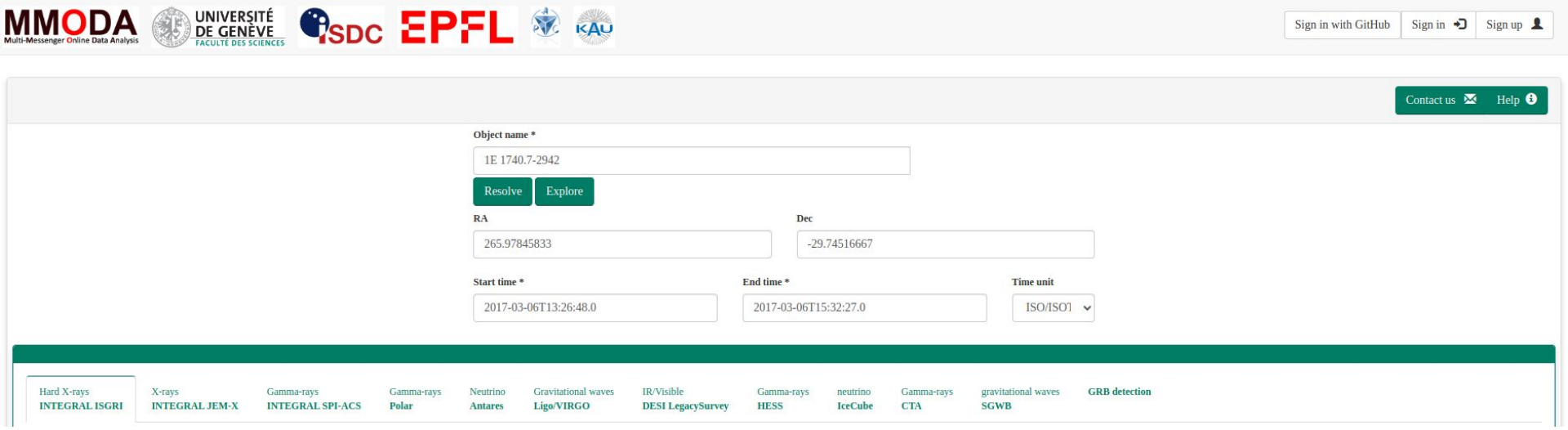
Multi-Messenger Online Data Analysis

https://www.astro.unige.ch/mmoda/

Andrii Neronov's Talk

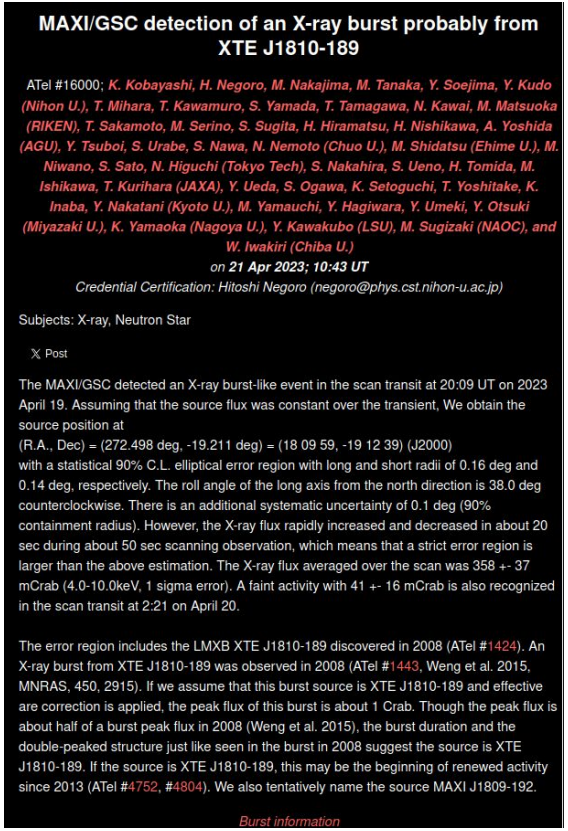
AstroORDAS

Astro Open Research Data Analysis Service



Goal

- A lot of information is written as and shared through Rapid Astronomical Bulletins such as ATels and GCNs
- **Automatise the analysis of ATels/GCNs**
- How ?
  - **Impact analysis**: study the past citations of ATels and GCNs
  - **Extract semantic content** from publications
  - **Find** and run **related workflows** by adding the processed data



MAXI/GSC detection of an X-ray burst probably from XTE J1810-189

ATel #16000; *K. Kobayashi, H. Negoro, M. Nakajima, M. Tanaka, Y. Soejima, Y. Kudo (Nihon U.), T. Mihara, T. Kawamuro, S. Yamada, T. Tamagawa, N. Kawai, M. Matsuoka (RIKEN), T. Sakamoto, M. Serino, S. Sugita, H. Hiramatsu, H. Nishikawa, A. Yoshida (AGU), Y. Tsuboi, S. Urabe, S. Nawa, N. Nemoto (Chuo U.), M. Shidatsu (Ehime U.), M. Niwano, S. Sato, N. Higuchi (Tokyo Tech), S. Nakahira, S. Ueno, H. Tomida, M. Ishikawa, T. Kurihara (JAXA), Y. Ueda, S. Ogawa, K. Setoguchi, T. Yoshitake, K. Inaba, Y. Nakatani (Kyoto U.), M. Yamauchi, Y. Hagiwara, Y. Umeki, Y. Otsuki (Miyazaki U.), K. Yamaoka (Nagoya U.), Y. Kawakubo (LSU), M. Sugizaki (NAOC), and W. Iwakiri (Chiba U.)*
*on **21 Apr 2023; 10:43 UT***
Credential Certification: Hitoshi Negoro (negoro@phys.cst.nihon-u.ac.jp)

Subjects: X-ray, Neutron Star

𝕏 Post

The MAXI/GSC detected an X-ray burst-like event in the scan transit at 20:09 UT on 2023 April 19. Assuming that the source flux was constant over the transient, We obtain the source position at
(R.A., Dec) = (272.498 deg, -19.211 deg) = (18 09 59, -19 12 39) (J2000)
with a statistical 90% C.L. elliptical error region with long and short radii of 0.16 deg and 0.14 deg, respectively. The roll angle of the long axis from the north direction is 38.0 deg counterclockwise. There is an additional systematic uncertainty of 0.1 deg (90% containment radius). However, the X-ray flux rapidly increased and decreased in about 20 sec during about 50 sec scanning observation, which means that a strict error region is larger than the above estimation. The X-ray flux averaged over the scan was 358 +- 37 mCrab (4.0-10.0keV, 1 sigma error). A faint activity with 41 +- 16 mCrab is also recognized in the scan transit at 2:21 on April 20.

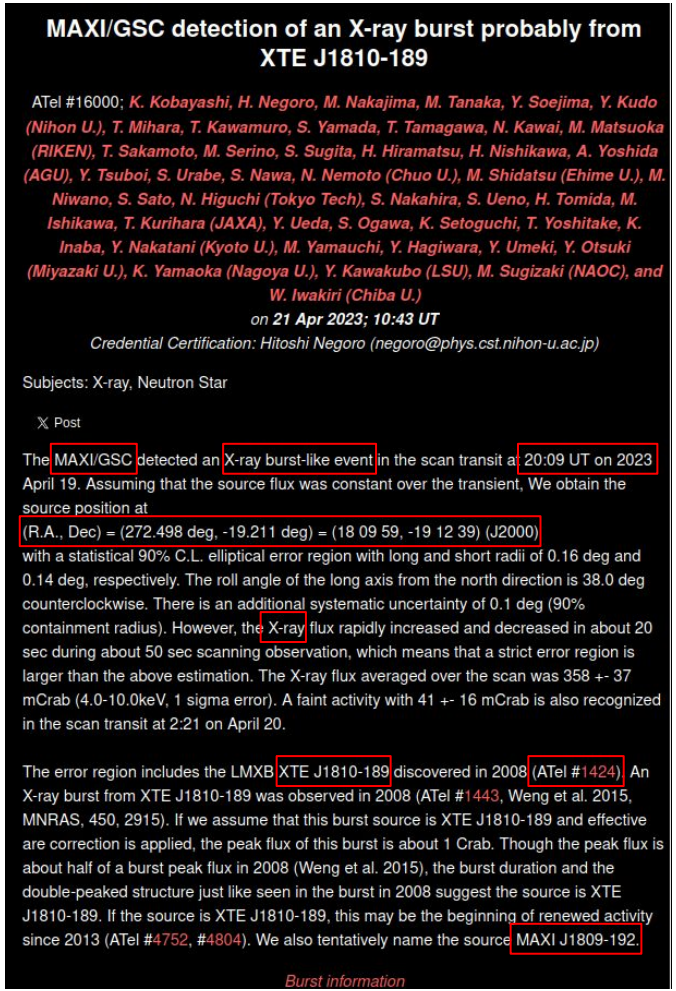
The error region includes the LMXB XTE J1810-189 discovered in 2008 (ATel #1424). An X-ray burst from XTE J1810-189 was observed in 2008 (ATel #1443, Weng et al. 2015, MNRAS, 450, 2915). If we assume that this burst source is XTE J1810-189 and effective are correction is applied, the peak flux of this burst is about 1 Crab. Though the peak flux is about half of a burst peak flux in 2008 (Weng et al. 2015), the burst duration and the double-peaked structure just like seen in the burst in 2008 suggest the source is XTE J1810-189. If the source is XTE J1810-189, this may be the beginning of renewed activity since 2013 (ATel #4752, #4804). We also tentatively name the source MAXI J1809-192.

Burst information

Sources of data

- ~16k ATels and ~32k GCNs.
  - on the 11th of Sept: Latest GCN 37454 and ATel 16811

- Include the references between ATels and GCN



MAXI/GSC detection of an X-ray burst probably from XTE J1810-189

ATel #16000; *K. Kobayashi, H. Negoro, M. Nakajima, M. Tanaka, Y. Soejima, Y. Kudo (Nihon U.), T. Mihara, T. Kawamuro, S. Yamada, T. Tamagawa, N. Kawai, M. Matsuoka (RIKEN), T. Sakamoto, M. Serino, S. Sugita, H. Hiramatsu, H. Nishikawa, A. Yoshida (AGU), Y. Tsuboi, S. Urabe, S. Nawa, N. Nemoto (Chuo U.), M. Shidatsu (Ehime U.), M. Niwano, S. Sato, N. Higuchi (Tokyo Tech), S. Nakahira, S. Ueno, H. Tomida, M. Ishikawa, T. Kurihara (JAXA), Y. Ueda, S. Ogawa, K. Setoguchi, T. Yoshitake, K. Inaba, Y. Nakatani (Kyoto U.), M. Yamauchi, Y. Hagiwara, Y. Umeki, Y. Otsuki (Miyazaki U.), K. Yamaoka (Nagoya U.), Y. Kawakubo (LSU), M. Sugizaki (NAOC), and W. Iwakiri (Chiba U.)*

*on **21 Apr 2023; 10:43 UT***

Credential Certification: Hitoshi Negoro (negoro@phys.cst.nihon-u.ac.jp)

Subjects: X-ray, Neutron Star

Post

The MAXI/GSC detected an X-ray burst-like event in the scan transit at 20:09 UT on 2023 April 19. Assuming that the source flux was constant over the transient, We obtain the source position at
(R.A., Dec) = (272.498 deg, -19.211 deg) = (18 09 59, -19 12 39) (J2000)
with a statistical 90% C.L. elliptical error region with long and short radii of 0.16 deg and 0.14 deg, respectively. The roll angle of the long axis from the north direction is 38.0 deg counterclockwise. There is an additional systematic uncertainty of 0.1 deg (90% containment radius). However, the X-ray flux rapidly increased and decreased in about 20 sec during about 50 sec scanning observation, which means that a strict error region is larger than the above estimation. The X-ray flux averaged over the scan was 358 +- 37 mCrab (4.0-10.0keV, 1 sigma error). A faint activity with 41 +- 16 mCrab is also recognized in the scan transit at 2:21 on April 20.

The error region includes the LMXB XTE J1810-189 discovered in 2008 (ATel #1424). An X-ray burst from XTE J1810-189 was observed in 2008 (ATel #1443, Weng et al. 2015, MNRAS, 450, 2915). If we assume that this burst source is XTE J1810-189 and effective are correction is applied, the peak flux of this burst is about 1 Crab. Though the peak flux is about half of a burst peak flux in 2008 (Weng et al. 2015), the burst duration and the double-peaked structure just like seen in the burst in 2008 suggest the source is XTE J1810-189. If the source is XTE J1810-189, this may be the beginning of renewed activity since 2013 (ATel #4752, #4804). We also tentatively name the source MAXI J1809-192.

Burst information

Data extraction

- Regular Expressions (REGEX)
  - rigorous, but not general

- Large Language Models (astroBERT, https://huggingface.co/adsabs/astroBERT):
  - Named Entity Recognition (NER)
  - generalisable, but may be less rigorous

Regular Expressions

- Source names (https://cds.unistra.fr/Dic/formats.html):
  - e.g. SDSS JHHMMSS.ss+DDMMSS.s (D-degree, H-hour, M-minute, S-second), ~40 patterns:
    - ~40k in ATels (~20k unique) and ~40k in GCNs (~12k unique)
  - Query: Simbad; TNS; FINK
  - Simbad
    - ~150 source classes and subclasses;
    - keep the 41 most interesting and common ones
- Telescope/Instrument/Observatory names (~780 entities):
  - built a RDF ontology (1) with: names, labels, types
    - GW, radio, IR, optical, UV, X-ray, γ-ray, cosmic-ray, neutrino
  - ~58k in ATels; ~110k in GCNs
- References between ATels and GCNs

- Work in Progress: R.A. - Dec. positions and Time

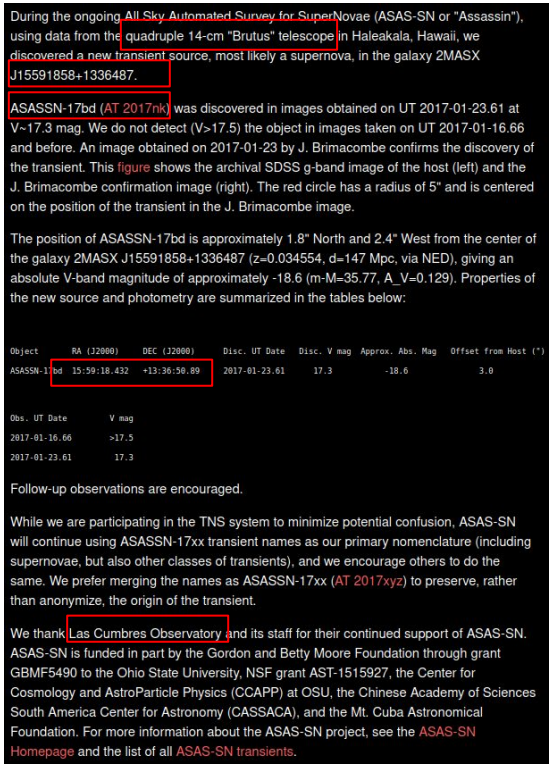
(1) https://gitlab.renkulab.io/andrei.variu/rule-based-source-and-telescope-detector/-/blob/master/data/public_ontologies/telescope_observatory_survey.ttl?ref_type=heads

astroBERT (NASA/ADS, https://arxiv.org/pdf/2112.00590)

- Based on Google's BERT LLM
- 110 mil params
- Trained/Validated/Tested on a dataset of ~400k articles:
  - ~4 bil. tokens, or ~3 bil. words and 121 mil. Sentences
- NER specific training:
  - dataset of ~1800 manually annotated sentences

astroBERT

	TEXT_ID	word	start	end	Score	Phrase	entity_group
0	ATel #10000	quadruple 14-cm	104	119	0.504776	During the ongoing All Sky Automated Survey fo...	Telescope
1	ATel #10000	Brutus	121	127	0.279041	During the ongoing All Sky Automated Survey fo...	Telescope
2	ATel #10000	telescope	129	138	0.910142	During the ongoing All Sky Automated Survey fo...	Telescope
3	ATel #10000	V	81	82	0.673355	ASASSN-17bd (AT 2017nk) was discovered in imag...	Wavelength
4	ATel #10000	V-band	177	183	0.818002	The position of ASASSN-17bd is approximately 1...	Wavelength
5	ATel #10000	g-band	36	42	0.802216	This figure shows the archival SDSS g-band ima...	Wavelength
6	ATel #10000	ASASSN-17bd (	0	13	0.614245	ASASSN-17bd (AT 2017nk) was discovered in imag...	CelestialObject
7	ATel #10000	2017nk	16	22	0.701770	ASASSN-17bd (AT 2017nk) was discovered in imag...	CelestialObject
8	ATel #10000	2MASX J15591858 + 1336487	238	261	0.779708	During the ongoing All Sky Automated Survey fo...	CelestialObject
9	ATel #10000	ASASSN-17bd 15	25	39	0.725085	Mag Offset from Host (") ASASSN-17bd 15:59:18....	CelestialObject
10	ATel #10000	432	46	49	0.698875	Mag Offset from Host (") ASASSN-17bd 15:59:18....	CelestialObject
11	ATel #10000	ASASSN-17bd	16	27	0.705332	The position of ASASSN-17bd is approximately 1...	CelestialObject
12	ATel #10000	2MASX J15591858+1336487	100	123	0.801807	The position of ASASSN-17bd is approximately 1...	CelestialObject
13	ATel #10000	ASASSN-17xx (AT 2017xyz	31	54	0.727070	We prefer merging the names as ASASSN-17xx (AT...	CelestialObject
14	ATel #10000	ASASSN-17xx	106	117	0.605917	While we are participating in the TNS system t...	CelestialObject
15	ATel #10000	:	39	40	0.395494	Mag Offset from Host (") ASASSN-17bd 15:59:18....	CelestialRegion
16	ATel #10000	:	42	43	0.779782	Mag Offset from Host (") ASASSN-17bd 15:59:18....	CelestialRegion
17	ATel #10000	+13:	50	54	0.667293	Mag Offset from Host (") ASASSN-17bd 15:59:18....	CelestialRegion
18	ATel #10000	:	56	57	0.772752	Mag Offset from Host (") ASASSN-17bd 15:59:18....	CelestialRegion
19	ATel #10000	Las Cumbres Observatory	9	32	0.886159	We thank Las Cumbres Observatory and its staff...	Observatory
20	ATel #10000	.	85	86	0.444411	We thank Las Cumbres Observatory and its staff...	Observatory
21	ATel #10000	ASAS - SN	0	7	0.854286	ASAS-SN is funded in part by the Gordon and Be...	Survey
22	ATel #10000	ongoing All Sky Automated Survey for SuperNova...	11	70	0.715511	During the ongoing All Sky Automated Survey fo...	Survey
23	ATel #10000	ASAS-SN	31	38	0.628564	For more information about the ASAS-SN project...	Survey
24	ATel #10000	ASAS-SN	56	63	0.400941	For more information about the ASAS-SN project...	Survey
25	ATel #10000	ASAS-SN	93	100	0.516507	For more information about the ASAS-SN project...	Survey
26	ATel #10000	SDSS	31	35	0.472629	This figure shows the archival SDSS g-band ima...	Survey
27	ATel #10000	ASAS - SN	78	85	0.504860	We thank Las Cumbres Observatory and its staff...	Survey
28	ATel #10000	ASAS - SN	78	85	0.824287	While we are participating in the TNS system t...	Survey



During the ongoing All Sky Automated Survey for SuperNovae (ASAS-SN or "Assassin"), using data from the quadruple 14-cm "Brutus" telescope in Haleakala, Hawaii, we discovered a new transient source, most likely a supernova, in the galaxy 2MASX J15591858+1336487.

ASASSN-17bd (AT 2017nk) was discovered in images obtained on UT 2017-01-23.61 at V~17.3 mag. We do not detect (V>17.5) the object in images taken on UT 2017-01-16.66 and before. An image obtained on 2017-01-23 by J. Brimacombe confirms the discovery of the transient. This figure shows the archival SDSS g-band image of the host (left) and the J. Brimacombe confirmation image (right). The red circle has a radius of 5" and is centered on the position of the transient in the J. Brimacombe image.

The position of ASASSN-17bd is approximately 1.8" North and 2.4" West from the center of the galaxy 2MASX J15591858+1336487 (z=0.034554, d=147 Mpc, via NED), giving an absolute V-band magnitude of approximately -18.6 (m-M=35.77, A_V=0.129). Properties of the new source and photometry are summarized in the tables below:

Object	RA (J2000)	DEC (J2000)	Disc. UT Date	Disc. V mag	Approx. Abs. Mag	Offset from Host (")
ASASSN-17bd	15:59:18.432	+13:36:50.89	2017-01-23.61	17.3	-18.6	3.0

Obs. UT Date	V mag
2017-01-16.66	>17.5
2017-01-23.61	17.3

Follow-up observations are encouraged.

While we are participating in the TNS system to minimize potential confusion, ASAS-SN will continue using ASASSN-17xx transient names as our primary nomenclature (including supernovae, but also other classes of transients), and we encourage others to do the same. We prefer merging the names as ASASSN-17xx (AT 2017xyz) to preserve, rather than anonymize, the origin of the transient.

We thank Las Cumbres Observatory and its staff for their continued support of ASAS-SN. ASAS-SN is funded in part by the Gordon and Betty Moore Foundation through grant GBMF5490 to the Ohio State University, NSF grant AST-1515927, the Center for Cosmology and AstroParticle Physics (CCAPP) at OSU, the Chinese Academy of Sciences South America Center for Astronomy (CASSACA), and the Mt. Cuba Astronomical Foundation. For more information about the ASAS-SN project, see the ASAS-SN Homepage and the list of all ASAS-SN transients.

- concatenate words separated by one character; still not sufficient (see “CelestialRegion”).
- “CelestialObject” entities are queried on Simbad
- “Wavelength” entities (“gamma-ray”, “x-ray”, etc) are further considered telescope-types

Subgoals

Given an ATel/GCN:

- can we predict whether it is followed-up by another ATel/GCN?
- can we predict the follow-up publication?
- we suppose that the follow-up ATel/GCN could represent an analysis performed by one of MMODA workflows.

Method (data)

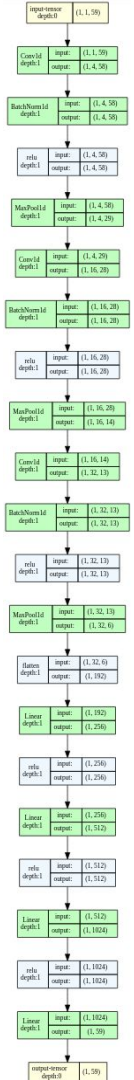
- Represent each text (ATel/GCN) as a vector with:
  - 41 source classes
  - 9 telescope types
  - 9 specific instruments, representing a workflow as well

- Data:
  - ~75k pairs (First ; Follow-up) of ATels/GCNs
  - ~48k labelled ATels/GCNs

Method (model)

- Model: convolutional neural-network (CNN, ~760k params)
  - similar models to predict the vector or the label

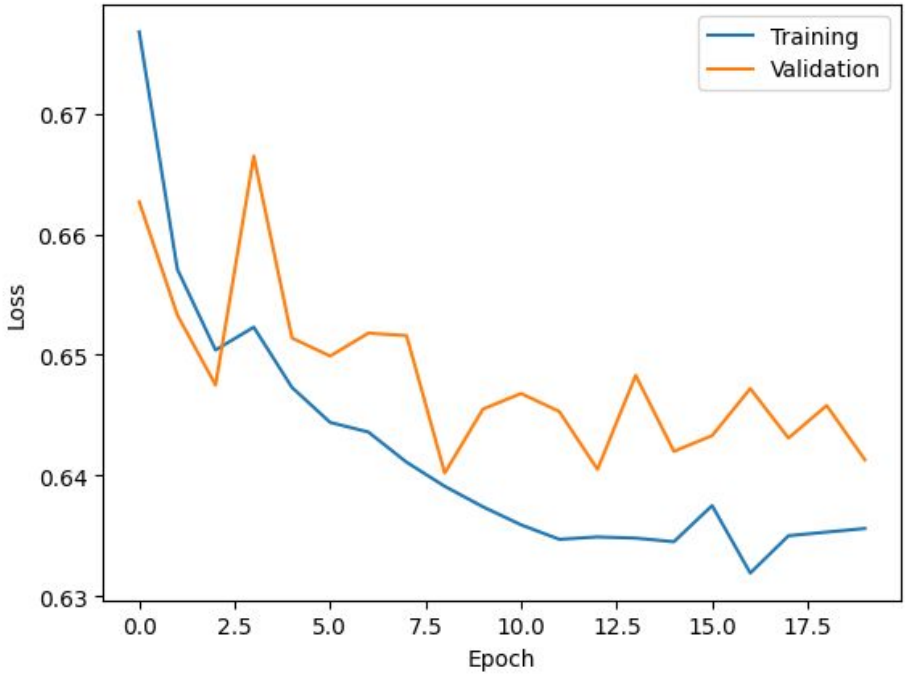
- Training:
  - vector-vector: 60E with a batch-size of 20, Mean Abs Error loss
  - vector-label: 20E with a batch-size of 20, Binary Cross Entropy loss



Results: Follow-up label prediction

Recall	Precision	Accuracy	F_1
0.32	0.69	0.61	0.43
TP/(FN+TP)	TP/(FP+TP)	(TP+TN) / (TP+FP+TN+FN)	

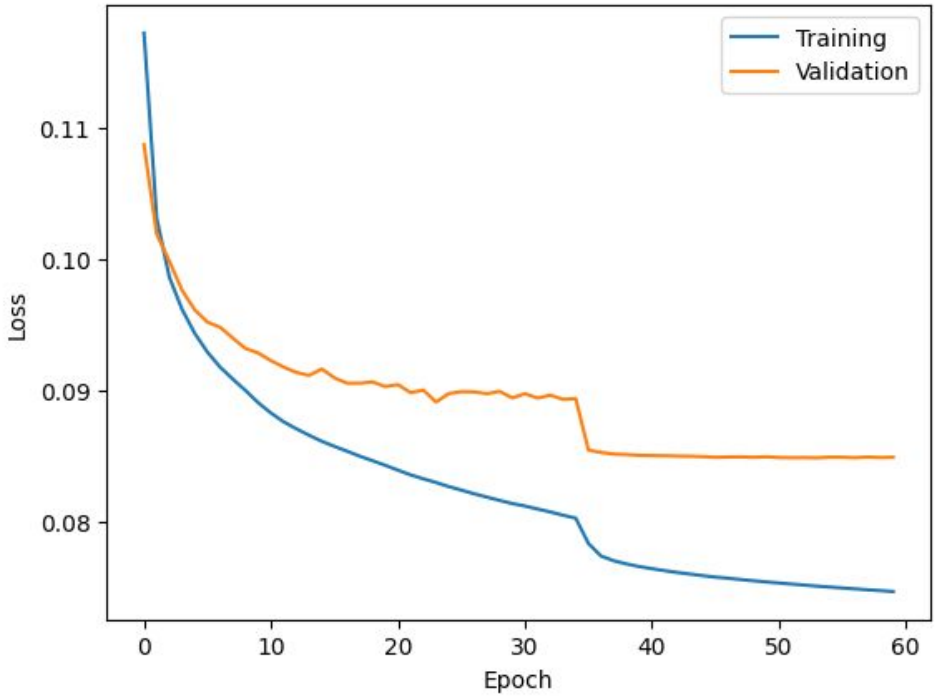
- ~50% of texts are followed-up
- CNN predicts only 15% of the time that they should be followed-up



Results: Vector prediction

	Recall	Precision	F_1
TH=0.01	0.76	0.76	0.76
TH=0.5	0.71	0.84	0.77

- Treat each value of the vector as an individual label
- Most predicted values are [0, 1] floats:
  - compute metrics for different thresholds
- We use a threshold of 0.01 for the next step



Results: Workflow linking

- Create a “URL” vector with the classes for each source from the initial text (ATel/GCN) and for each workflow
- Normalize this vector and the predicted follow-up vector
- Compute the dot product between the two vectors and use it as a score

	Legend	ATel #16802	Follow-up	URL_0PSR J1841-0456
0	MIR	0	0.000000	0.0
1	rB	0	0.000000	0.0
2	BLL	0	0.000000	0.0
3	UV	0	0.000000	0.0
4	LIN	0	0.000000	0.0
5	SN*	0	0.000000	0.0
6	HI	0	0.000000	0.0
7	rG	0	0.000000	0.0
8	SetGalaxies	0	0.000000	0.0
9	GWE	0	0.000000	0.0
10	Bla	0	0.000000	0.0
11	XB*	0	0.000000	0.0
12	NIR	0	0.000000	0.0
13	CV*	0	0.000000	0.0
14	SN?	0	0.000000	0.0
15	Weird	0	0.000000	0.0
16	gam	1	1.121320	1.0
17	Q?	0	0.000000	0.0
18	SyG	0	0.000000	0.0
19	CV?	0	0.000000	0.0
20	cm	0	0.000000	0.0
21	ev	0	0.000000	0.0
22	*	1	0.719775	1.0
23	Opt	0	0.000000	0.0
24	gB	0	0.000000	0.0
25	var	0	0.000000	0.0
26	V*	0	0.000000	0.0
27	**	0	0.000000	0.0
28	Cl*	0	0.000000	0.0
29	IR	0	0.000000	0.0
30	AG?	0	0.000000	0.0
31	QSO	0	0.000000	0.0
32	G	0	0.000000	0.0
33	Bz?	0	0.000000	0.0
34	X	1	0.942385	1.0
35	grv	0	0.000000	0.0
36	ISM	0	0.000000	0.0
37	As*	0	0.000000	0.0
38	Rad	1	1.022737	1.0
39	AGN	0	0.000000	0.0
40	BL?	0	0.000000	0.0
41	gamma-ray	1	1.008096	1.0
42	x-ray	3	1.847898	0.0
43	ultraviolet	0	0.000000	0.0
44	optical	0	0.052257	0.0
45	infrared	0	0.000000	0.0
46	radio	0	0.301226	0.0
47	cosmic-ray	0	0.000000	0.0
48	gravitational-wave	0	0.000000	0.0
49	neutrino	0	0.000000	0.0
50	INTEGRAL	0	0.000000	1.0
51	ISGRI	0	0.000000	0.0
52	JEM-X	0	0.000000	0.0
53	SPI-ACS	0	0.000000	1.0
54	ANTARES	0	0.000000	0.0
55	LIGO/VIRGO	0	0.000000	0.0
56	IceCube	0	0.000000	0.0
57	HESS	0	0.000000	0.0
58	CTA/CTAO	0	0.000000	0.0

Results: Workflow linking

ATel #16802: NICER, NuSTAR, and Swift-XRT observations of the Magnetar 1E 1841-045

Source: table_vectorized_url_scores [2024.09.12T14:11:40]

Download

Show 50 entries Search:

URL Name	Scores	URL
URL_8PSR J1841-0456	0.116937854886055	https://www.astro.unige.ch/mmoda/?instrument=polar&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_8PSR J1841-0456	0.116937854886055	https://www.astro.unige.ch/mmoda/?instrument=grb_detection&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_3PSR J1841-0456	0.09809737971850804	https://www.astro.unige.ch/mmoda/?instrument=isgri&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_4PSR J1841-0456	0.09809737971850804	https://www.astro.unige.ch/mmoda/?instrument=jemx&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_1PSR J1841-0456	0.09744821240504581	https://www.astro.unige.ch/mmoda/?instrument=cta&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_2PSR J1841-0456	0.09744821240504581	https://www.astro.unige.ch/mmoda/?instrument=hess&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_9PSR J1841-0456	0.09372086245566608	https://www.astro.unige.ch/mmoda/?instrument=legacysurvey&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_0PSR J1841-0456	0.083527039204325	https://www.astro.unige.ch/mmoda/?instrument=spi_acs&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_5PSR J1841-0456	0.0770429695645968	https://www.astro.unige.ch/mmoda/?instrument=gw&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_6PSR J1841-0456	0.0770429695645968	https://www.astro.unige.ch/mmoda/?instrument=icecube&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006
URL_7PSR J1841-0456	0.0770429695645968	https://www.astro.unige.ch/mmoda/?instrument=antares&src_name=PSR+J1841-0456&RA=280.33037499999995&DEC=-4.9365000000000006

Conclusion

- A first attempt to automatise this analysis process
- The suggested workflows seem to be what an astrophysicist would suggest as well
- We still want to check how we can use GPT3.5 (or other LLM)
- Need to improve the CNN prediction: better representation/data or better model
- Q: Could it be interesting to perform similar analysis on other type of alerts?