ALeRCE
Automatic Learning for the
Rapid Classification of Events

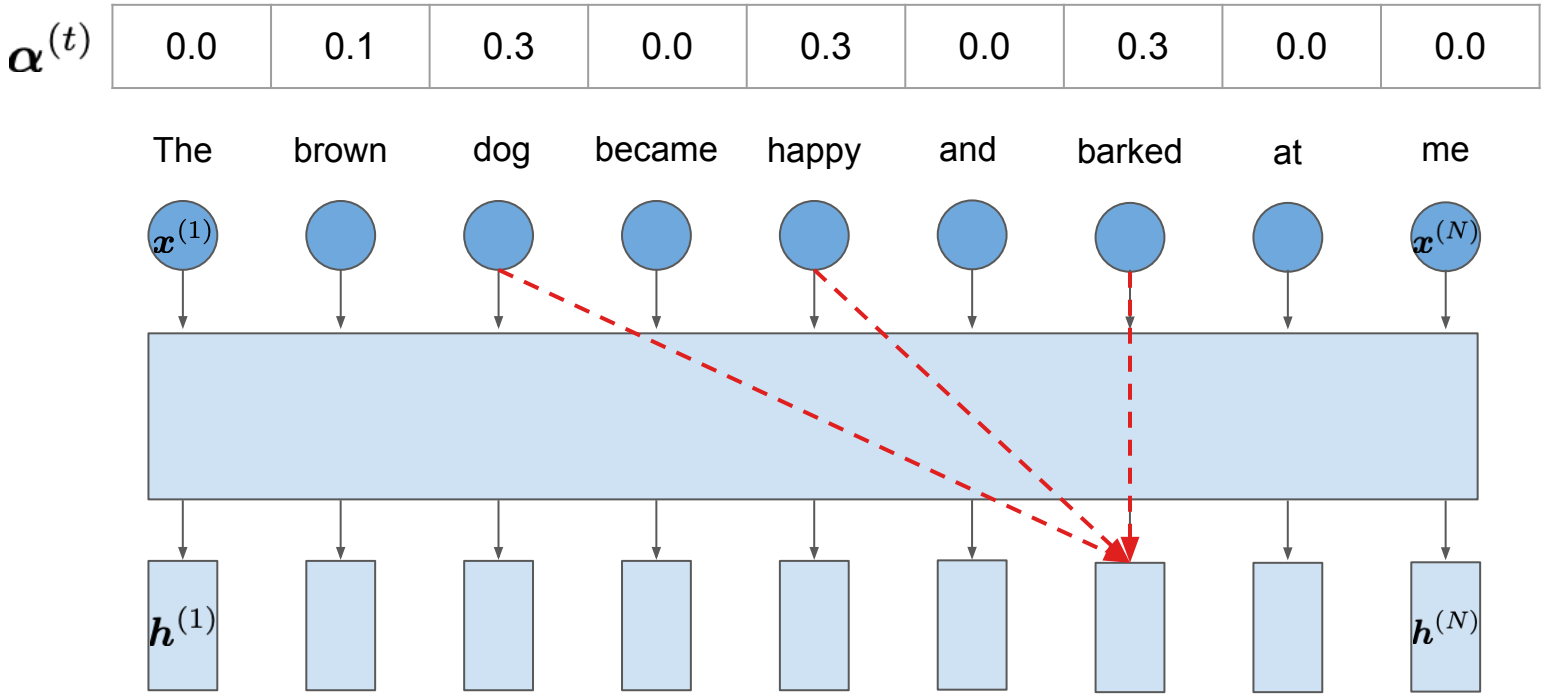
The Arrival of the Transformers for Advanced Analysis of Alert Streams

Guillermo Cabrera-Vives
Universidad de Concepción
Millenium Institute of Astrophysics
Millennium Nucleus on Young Exoplanets and their Moons

F. Förster; G. Cabrera-Vives; E. Castillo-Navarrete; P. A. Estévez; P. Sánchez-Sáez; J. Arredondo; F. E. Bauer; A. Bayo; R. Carrasco-Davis; M. Catelan; F. Elorrieta; S. Eyheramendy; P. Huijse; G. Pignata; E. Reyes; I. Reyes; D. Rodríguez-Mancini; D. Ruz-Mieres; C. Valenzuela; I. Álvarez-Maldonado; N. Astorga; J. Borissova; A. Clocchiatti; D. De Cicco; C. Donoso-Oliva; M. J. Graham; L. Hernández-García; R. Kurtev; A. Mahabal; J.C. Maureira; R. Molina-Ferreiro; A. Moya; A. Muñoz Arancibia; W. Palma; M. Pérez-Carrasco; A. Papageorgiou; P. Protopapas; M. Romero; L. Sabatini-Gacitua; A. Sánchez; J. San Martín; C. Sepúlveda-Cobo; E. Vera; J. R. Vergara

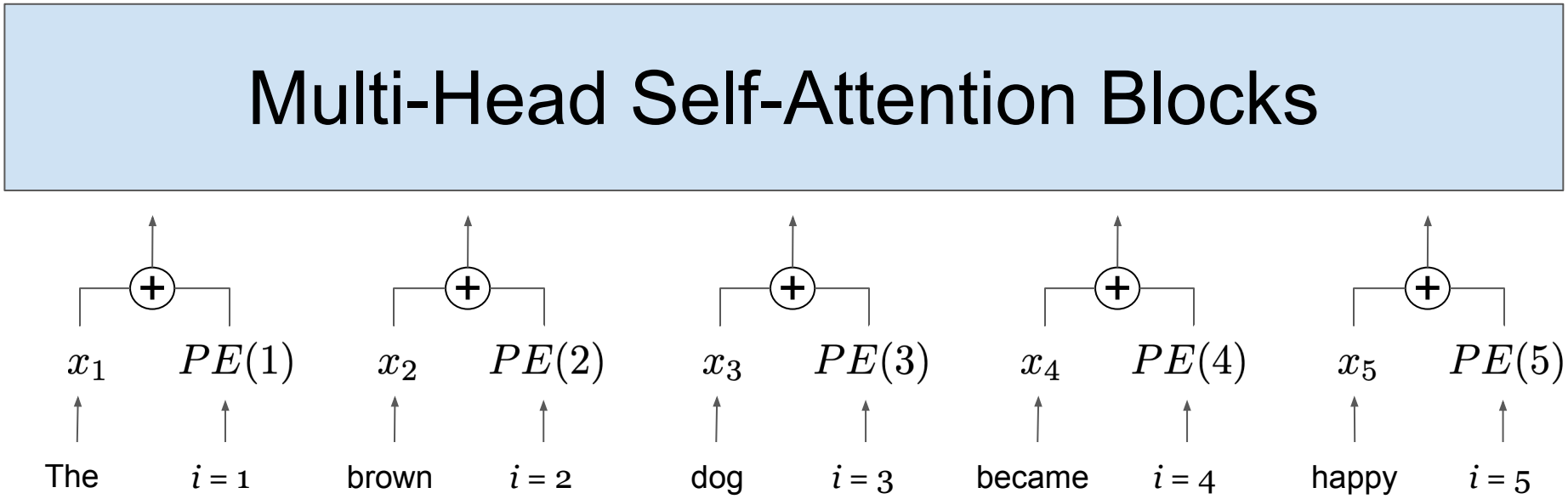
What is a transformer?

Self-attention

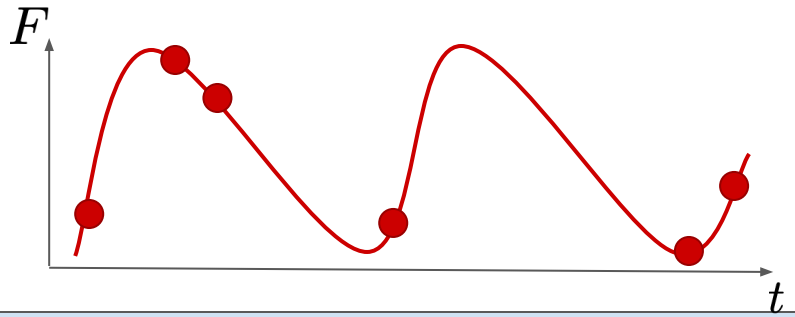


Transformers in NLP

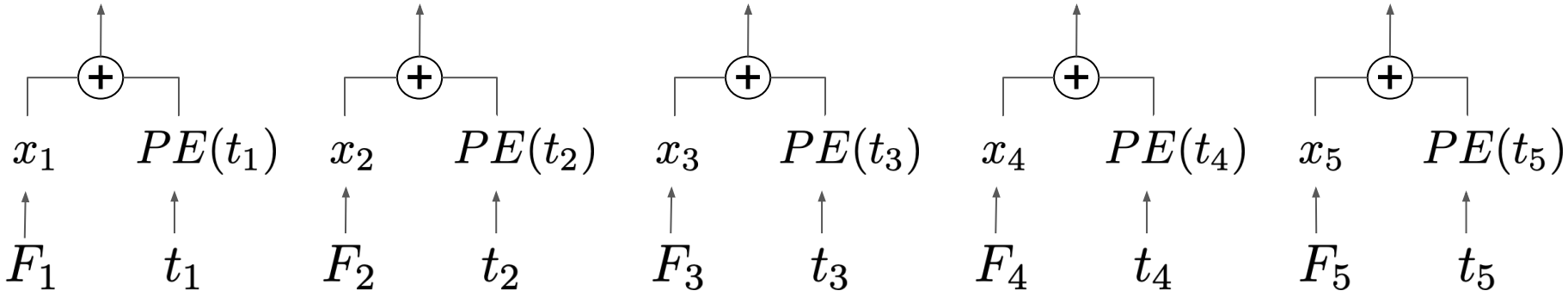
“The brown dog became happy and barked at me”



Transformers for Light Curves

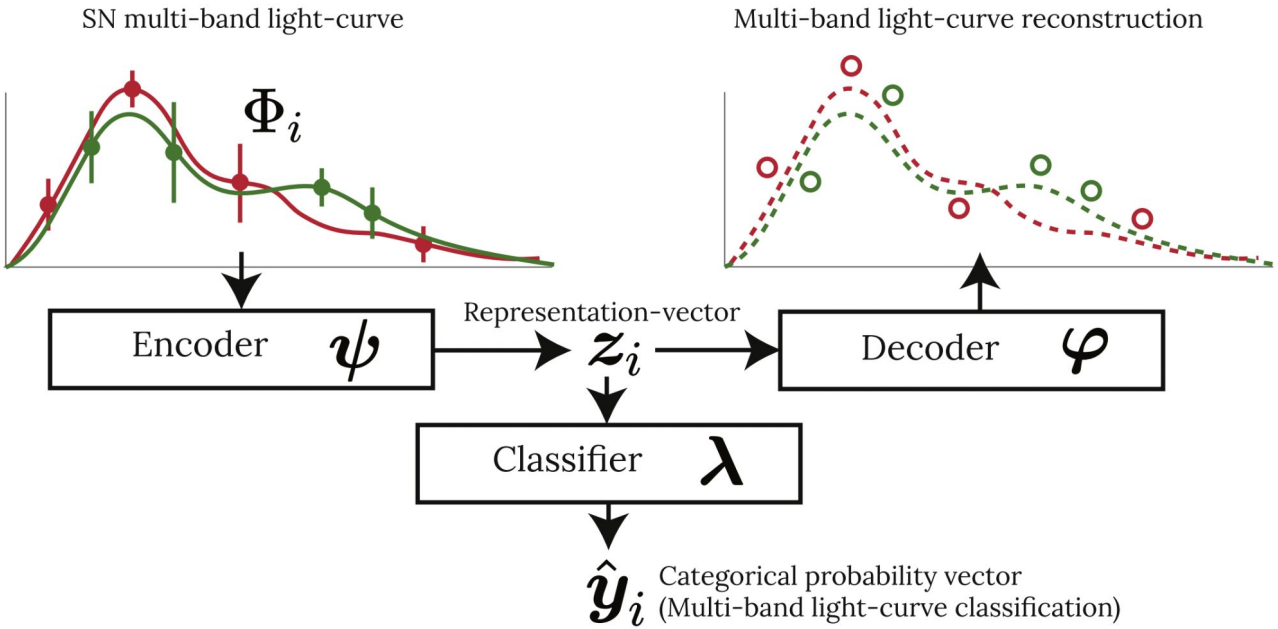


Multi-Head Self-Attention Blocks



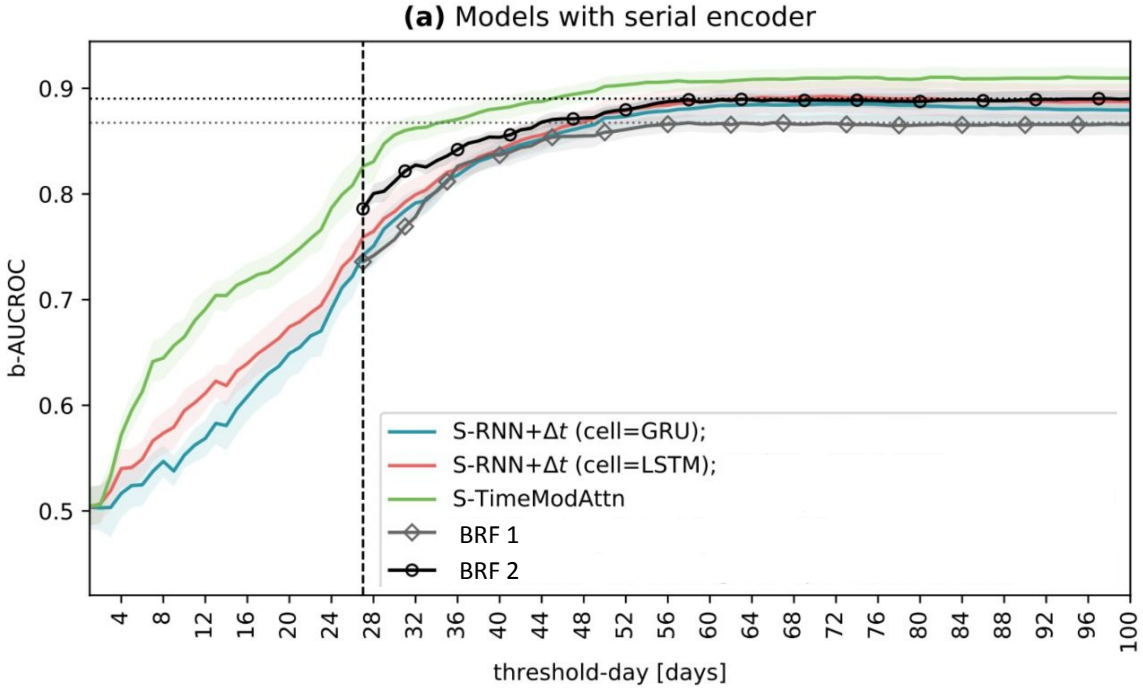

ALeRCE
Automatic Learning for the
Rapid Classification of Events
Transformers for SNe
Classification

Multiband Supernova Classification

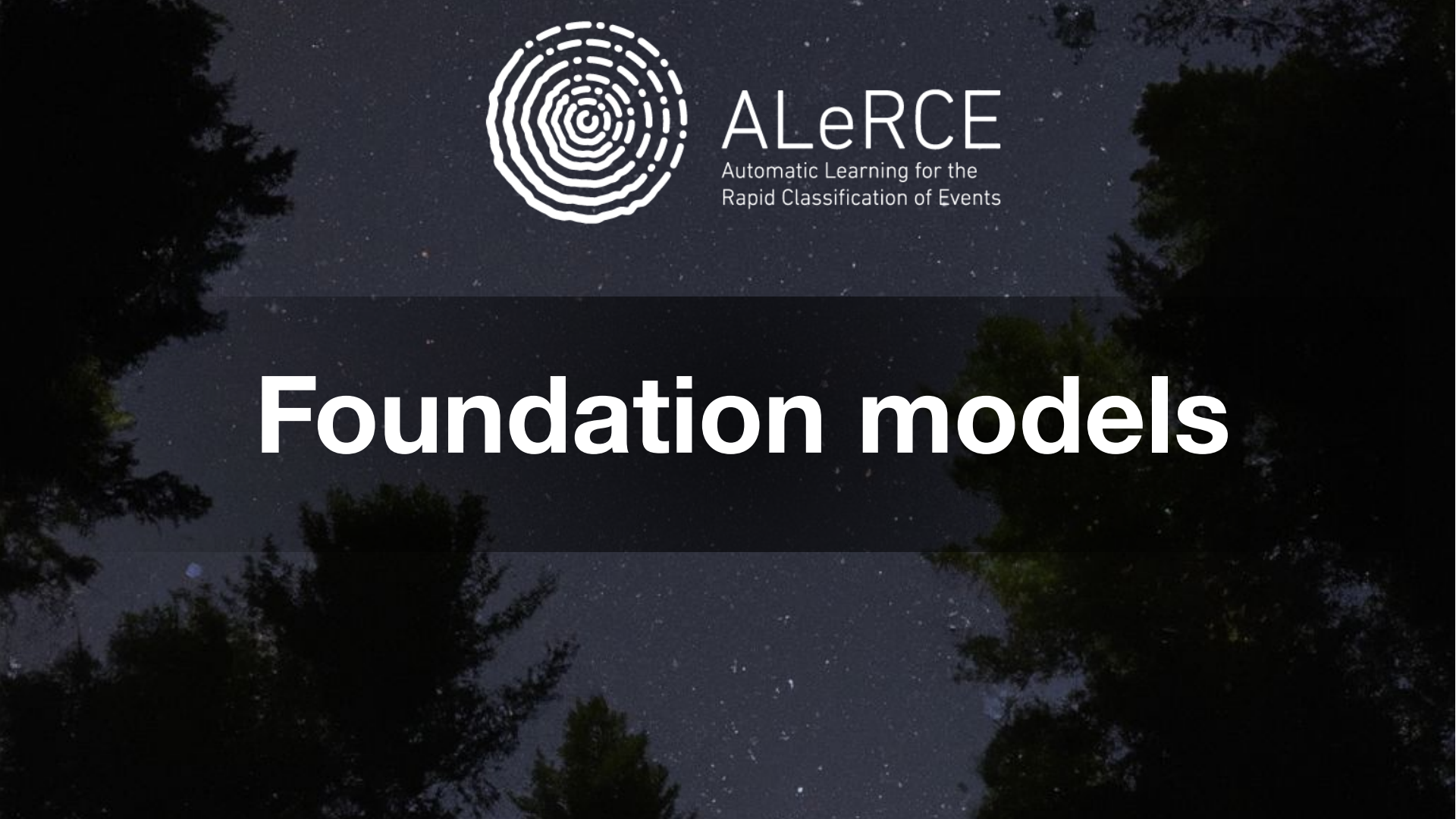


Pimentel O.. et.al., Deep Attention-based Supernovae Classification of Multiband Light Curves, AJ, 2023

Multiband Supernova Classification



Pimentel O.. et.al., Deep Attention-based Supernovae Classification of Multiband Light Curves, AJ, 2023


ALeRCE
Automatic Learning for the
Rapid Classification of Events
Foundation models

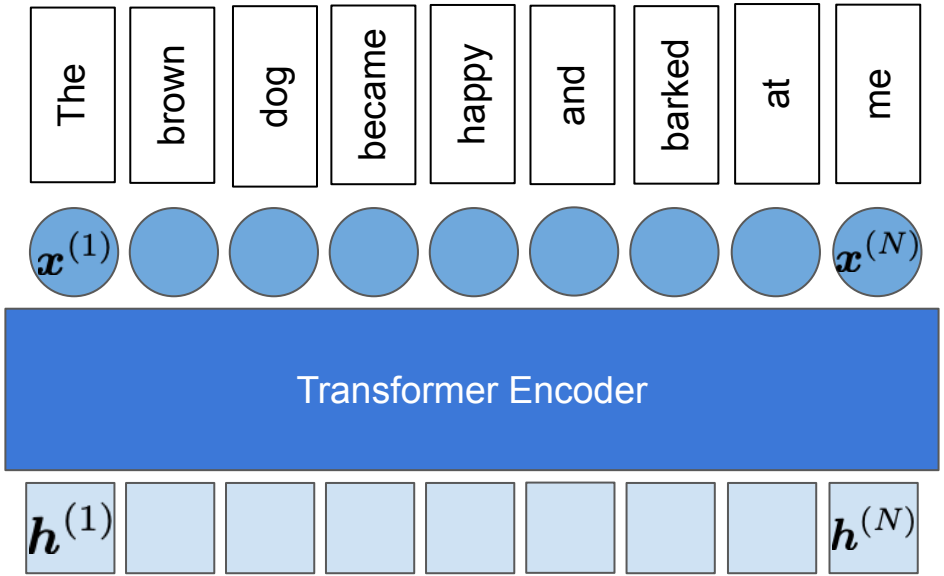
Foundation Models

Foundation models (e.g., BERT, GPT-3, CLIP, Codex) are models trained on broad data at scale such that they can be adapted to a wide range of downstream tasks.

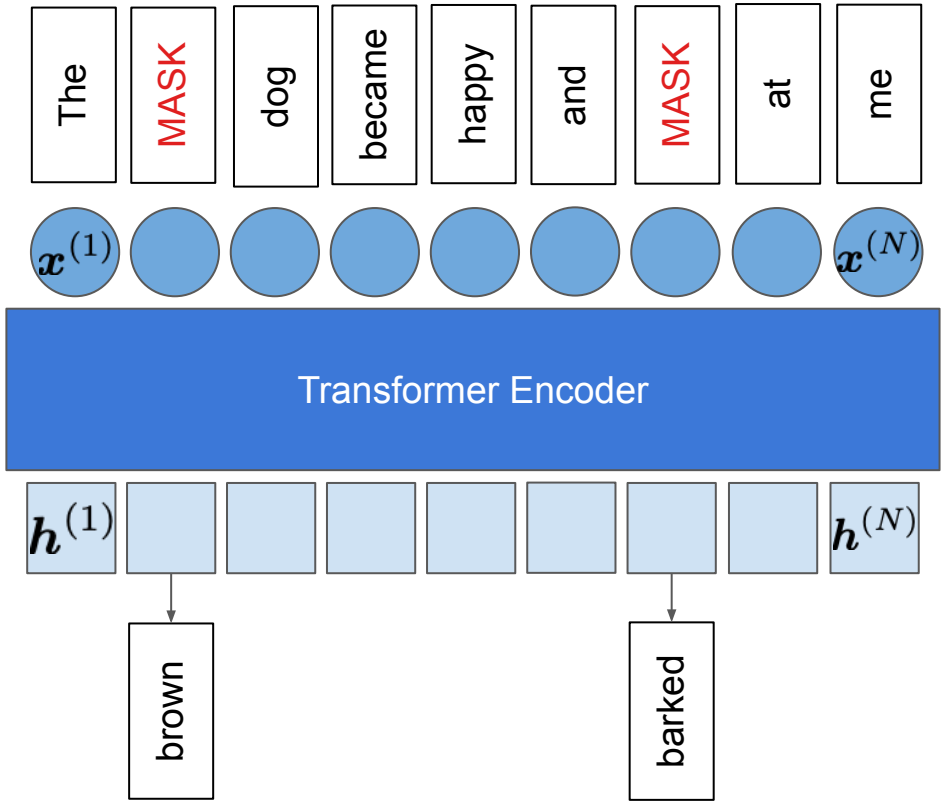
https://hai.stanford.edu/news/introducing-center-research-foundation-models-crfm

Devlin et.al., BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, 2019, NAACL-HLT

BERT


The
brown
dog
became
happy
and
barked
at
me
$x^{(1)}$
$x^{(N)}$
Transformer Encoder
$h^{(1)}$
$h^{(N)}$

BERT



BERT masks 15% of all tokens in each sequence at random.

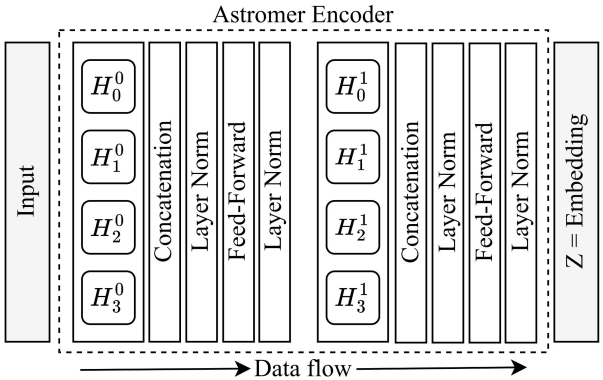
Masked Language Model (MLM)

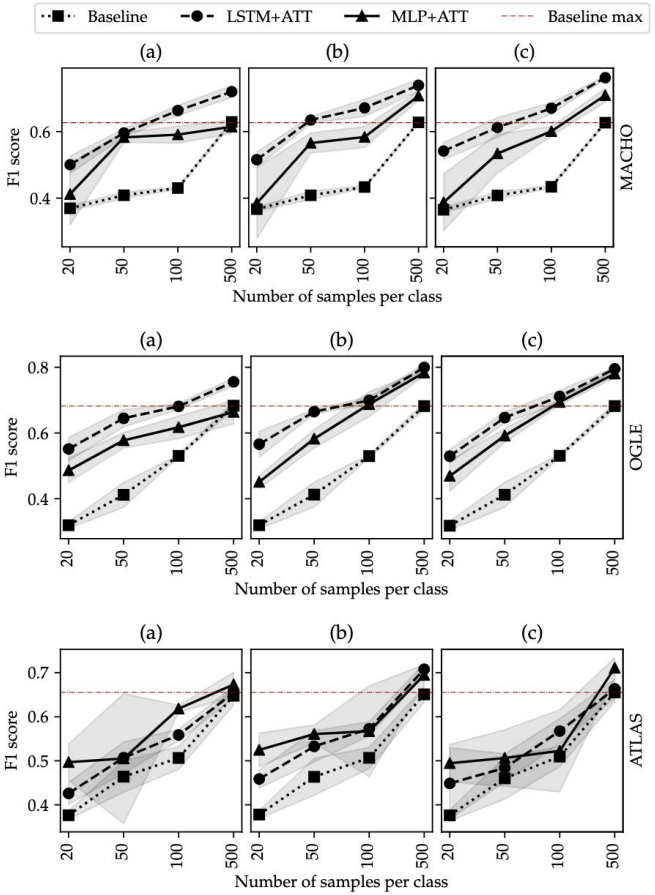
Prediction (softmax)

ASTROMER

Pretraining

MACHO unlabeled;
Alcock et.al., 2000
1,529,386 R-band light curves

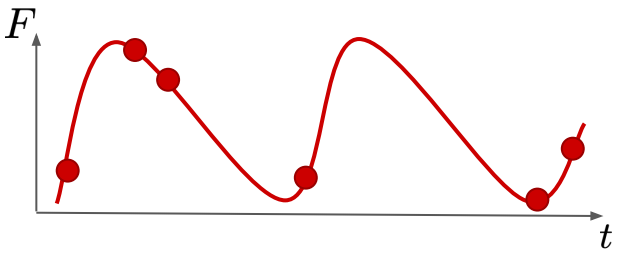


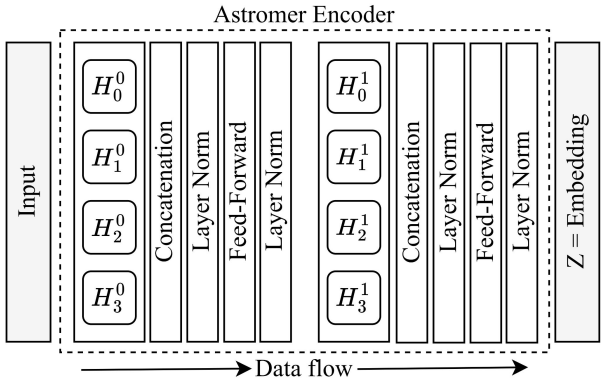


Donoso-Oliva et.al., ASTROMER: A transformer-based embedding for the representation of light curves, 2023, A&A

Positional Encodings for Light Curve Transformers

- We test different PEs for light curve transformers
- We evaluate their performances in terms of:
  - reconstruction of astronomical time series
  - classification of variable stars





D. Moreno-Cartagena et.al., Positional Encodings for Light Curve Transformers: Playing with Positions and Attention, ICML ML4Astro 2023

Positional Encodings for Light Curve Transformers

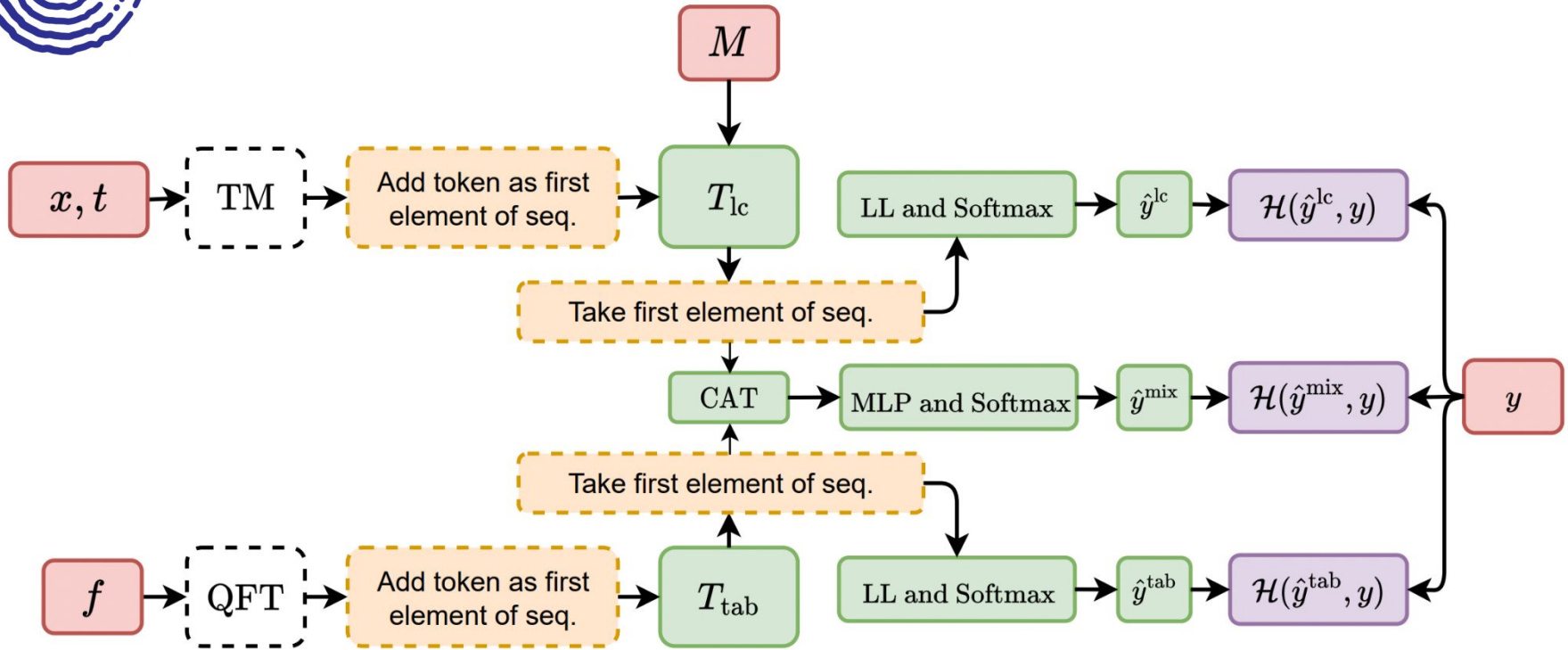
PE type	MACHO unlab.		MACHO lab.				OGLE	ATLAS
			Full	3/4	1/2	1/4		
	RMSE	Time (epochs)	F1 (%)	F1 (%)	F1 (%)	F1 (%)	F1 (%)	F1 (%)
Baseline	.170	6d 14h (523)	71.6 ± 1.9	69.2 ± 1.9	66.2 ± 1.9	63.3 ± 1.5	71.3 ± 1.1	65.8 ± 1.4
Trainable	**.169**	2d 13h (202)	72.9 ± 2.1	72.3 ± 1.0	$\mathbf{71.0 \pm 1.0}$	$\mathbf{69.0 \pm 0.5}$	74.9 ± 1.4	65.4 ± 1.8
Fourier	.170	1d 20h (142)	73.0 ± 1.1	70.2 ± 1.9	67.8 ± 0.9	62.9 ± 2.0	72.0 ± 0.8	$\mathbf{69.6 \pm 0.1}$
Recurrent	.197	0d 16h (048)	67.1 ± 1.8	63.5 ± 2.5	59.7 ± 1.9	54.6 ± 1.3	70.7 ± 1.1	68.3 ± 0.9
Tupe-A	.219	0d 17h (084)	67.3 ± 1.6	66.1 ± 1.4	64.9 ± 1.0	60.8 ± 0.9	71.0 ± 1.0	67.5 ± 0.9
Concat	.170	3d 01h (237)	$\mathbf{73.4 \pm 1.1}$	$\mathbf{73.1 \pm 1.7}$	70.9 ± 1.7	$\mathbf{69.0 \pm 1.8}$	74.5 ± 1.3	68.1 ± 0.6
PEA	.199	0d 17h (058)	69.7 ± 0.9	68.9 ± 1.8	68.0 ± 1.0	65.5 ± 2.5	$\mathbf{76.3 \pm 1.2}$	66.9 ± 1.0

- Trainable PEs (added or concatenated) suffer less degradation with sparser light curves creating a better representation of temporal information

- PEA outperforms other models for OGLE, while Fourier achieves the best performance over ATLAS

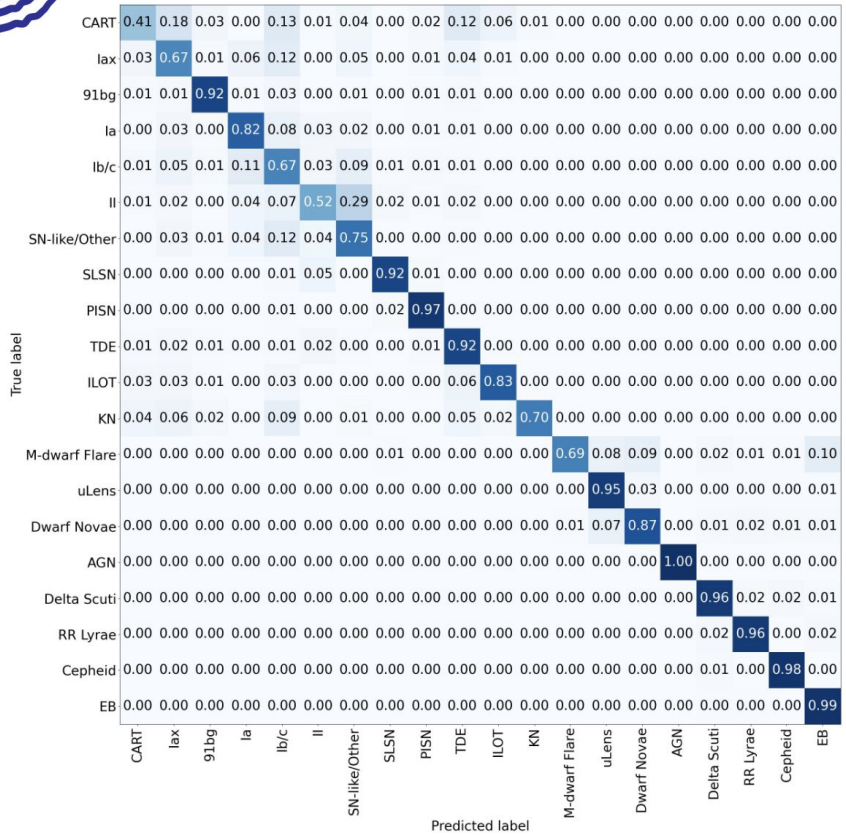

ALeRCE
Automatic Learning for the
Rapid Classification of Events
Astronomical Transformer for
time series And Tabular data

ATAT



Astorga et.al., ATAT: Astronomical Transformer for time series And Tabular data, DOI: 10.21203/rs.3.rs-2395110/v1

ATAT


CART 0.41 0.18 0.03 0.00 0.13 0.01 0.04 0.00 0.02 0.12 0.06 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Iax 0.03 0.67 0.01 0.06 0.12 0.00 0.05 0.00 0.01 0.04 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
91bg 0.01 0.01 0.92 0.01 0.03 0.00 0.01 0.00 0.01 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Ia 0.00 0.03 0.00 0.82 0.08 0.03 0.02 0.00 0.01 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
Ib/c 0.01 0.05 0.01 0.11 0.67 0.03 0.09 0.01 0.01 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
II 0.01 0.02 0.00 0.04 0.07 0.52 0.29 0.02 0.01 0.02 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
SN-like/Other 0.00 0.03 0.01 0.04 0.12 0.04 0.75 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
SLSN 0.00 0.00 0.00 0.00 0.01 0.05 0.00 0.92 0.01 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
PISN 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.02 0.97 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
TDE 0.01 0.02 0.01 0.00 0.01 0.02 0.00 0.00 0.01 0.92 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
ILOT 0.03 0.03 0.01 0.00 0.03 0.00 0.00 0.00 0.00 0.06 0.83 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
KN 0.04 0.06 0.02 0.00 0.09 0.00 0.01 0.00 0.00 0.05 0.02 0.70 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00
M-dwarf Flare 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.00 0.00 0.00 0.00 0.69 0.08 0.09 0.00 0.02 0.01 0.01 0.10
uLens 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.95 0.03 0.00 0.00 0.00 0.00 0.01
Dwarf Novae 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.07 0.87 0.00 0.01 0.02 0.01 0.01
AGN 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 1.00 0.00 0.00 0.00 0.00
Delta Scuti 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.96 0.02 0.02 0.01
RR Lyrae 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.02 0.96 0.00 0.02
Cepheid 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.01 0.00 0.98 0.00
EB 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.99
True label
CART Iax 91bg Ia Ib/c II SN-like/Other SLSN PISN TDE ILOT KN M-dwarf Flare uLens Dwarf Novae AGN Delta Scuti RR Lyrae Cepheid EB
Predicted label

Takeaways

- Transformers are becoming the state-of-the-art in many fields, including the analysis of alert streams.
- Foundation models for light curves are becoming a reality.
- The future: multi-modal, multi-stream, multi-task generalist models for astronomy?

ALeRCE

Automatic Learning for the Rapid Classification of Events

The Arrival of the Transformers for Advanced Analysis of Alert Streams

Guillermo Cabrera-Vives
Universidad de Concepción
Millenium Institute of Astrophysics
Millennium Nucleus on Young Exoplanets and their Moons

guillecabrera@inf.udec.cl