Predicting the Age of Astronomical Transients from Real-Time Multivariate Time Series

Hali Huang[1]   **Daniel Muthukrishna**[1*]   **Prajna Nair**[1]   **Zimi Zhang**[1]
Michael Fausnaugh[1,2]   **Torsha Majumder**[1]   **Ryan J. Foley**[3]   **George R. Ricker**[1]
[1]MIT   [2]Texas Tech University   [3]UC Santa Cruz
danmuth@mit.edu

Abstract

Astronomical transients, such as supernovae and other rare stellar explosions, have been instrumental in some of the most significant discoveries in astronomy. New astronomical sky surveys will soon record unprecedented numbers of transients as sparsely and irregularly sampled multivariate time series. To improve our understanding of the physical mechanisms of transients and their progenitor systems, early-time measurements are necessary. Prioritizing the follow-up of transients based on their age along with their class is crucial for new surveys. To meet this demand, we present the first method of predicting the age of transients in real-time from multi-wavelength time-series observations. We build a Bayesian probabilistic recurrent neural network. Our method can accurately predict the age of a transient with robust uncertainties as soon as it is initially triggered by a survey telescope. This work will be essential for the advancement of our understanding of the numerous young transients being detected by ongoing and upcoming astronomical surveys.

1 Introduction

The study of astronomical transients is crucial for understanding the nature of the universe, from one class of supernovae being a fundamental probe for measuring the expansion rate of the universe, to probing nucleosynthesis in extreme conditions. New astronomical sky surveys such as the Vera Rubin Observatory's Legacy Survey of Space and Time (LSST, Ivezić et al. 7) are expected to observe over 10 million transient alerts each night, increasing the discovery rate of known supernovae by 100 times. Given these massive data volumes and the limited number of worldwide follow-up observatories, only a tiny fraction of these events can expect to receive detailed spectroscopic or multi-wavelength follow-up observations.

In recent years, several efforts have been made to automate the identification of transients using machine learning techniques. However, most previous approaches require the full phase coverage of the transient, limiting the scientific questions that can be answered about these events. Early-time follow-up observations, including high-cadence photometry and time-resolved spectroscopy shortly after a transient's explosion, are necessary for gaining insights into the progenitor systems driving the event and advancing our understanding of the transient's physical mechanisms [9]. This has motivated the need for the early identification of young transients (such as the Young Supernova Experiment [8]), which is becoming increasingly critical with the volume of data generated by new survey telescopes [e.g. 11, 12, 17, 13, 1]. However, early classification on real data has not yet been effectively demonstrated in the literature. Therefore, we propose that accurately measuring a transient's age is a crucial criterion for selecting which transients to follow up from the abundance of new data. In this paper, we present the first method for real-time identification of a transient's age, enabling the prioritization of young transients and facilitating rapid follow-up and detailed observations of their progenitor systems. Using this framework alongside transient classifiers and anomaly detection frameworks will be essential for making discoveries in new large-scale surveys.

2 Data

To evaluate our method, we collected a set of labeled transients that have been observed by the Zwicky Transient Facility (ZTF, [2]) and the Transiting Exoplanet Survey Satellite (TESS, [14]). Our dataset comprises 805 transients, each represented by a multi-channel time series known as a *light curve*.

The majority of transients, 529 of them, in our dataset are unclassified. The classified transients consist of 230 Type Ia Supernovae, 37 core collapse supernovae and 9 other types. Type Ia supernovae result from the thermonuclear explosion of a binary system with a carbon-oxygen white dwarf accreting matter from a companion star. Core collapse supernovae include Type Ibc (SNIbc) and Type II (SNII) supernovae and are the result of the core collapse of massive stars. The other category consists of a variety of classifications found uncommonly in the dataset such as Tidal Disruption Events (TDE), variable stars (CVs), and Type I Superluminous supernovae (SLSN).

These light curves include flux and flux uncertainty measurements in the ZTF-g (green wavelength range), ZTF-r (red wavelength range), and TESS (red to infrared wavelength range) passbands. Notably, the light curves are sparsely and irregularly sampled. The ZTF passbands exhibit a mean cadence of 3-day intervals, while the TESS bands have a mean cadence of 10 minutes but cover only a short duration (~ 27 days) of the light curve. An example light curve is depicted in Fig. 1a.

We define the *trigger* as the time when ZTF first detects a 5σ change in the flux. We exclude observations more than 30 days before the trigger and more than 70 days after the trigger to focus on the transient phase of the light curves. We used the forced photometry service from ZTF [10] to quantify the non-detection observations early in a transient light curve. These early measurements contain crucial information about the host galaxy that has been shown to be very useful for transient identification and modeling [e.g. 4, 5].

3 Method

The goal of this work is to predict the age of a transient based on photometric time-series data. The age of a transient is typically defined by either the number of days after maximum light or the number of days after first light, which is approximately coincident with the supernova explosion time. We have built separate models to predict the time of first light t_0 and the time of maximum light $t_{\max}$. It is important that the age can be identified as early as possible - ideally within a few days after the time of trigger.

Dealing with missing data is a major challenge in time-series machine learning, with most works opting for some sort of interpolation of the data. However, these interpolation methods can introduce out-of-distribution data sets that are not representative of real data. In our work, we bypass the missing data problem by explicitly providing the neural network with time t relative to trigger for each data point. We set each column of the input matrix representing the multi-channel time-series data set as follows,

$$\boldsymbol{X}_{spt} = [t, \lambda_p, D_{spt}, \sigma_{D,spt}]^\top, \tag{1}$$

where D_{spt} is the observed flux for transient s, passband p, at time t, $\sigma_{D,spt}$ is the corresponding flux uncertainty, and λ_p is the central wavelength of each passband in units of 10^{-7}m as follows $\{\lambda_g = 4.767, \lambda_r = 6.215, \lambda_{\text{TESS}} = 7.865\}$. Hence, each column of the matrix represents an observation with the flux and uncertainty at a particular time and passband.

3.1 Obtaining Ground Truth Labels

In order to apply supervised learning methods to our dataset, we require ground truth labels for the ages of each transient. Because of missing light curve data, the ground truth is unknown, and we need to estimate the times of maximum and first light. We developed a Bayesian parametric model to estimate the times and their uncertainties. The training labels for the time of maximum light were calculated by fitting the full light curves with the Supernova Parametric Model (SPM [15]), which is as follows:

$$F = \frac{A(1 - \beta'(\frac{t-t_0}{t_1-t_0}))}{1 + e^{-\frac{t-t_0}{\tau_{\text{rise}}}}} \cdot \left[1 - \sigma\left(\frac{t-t_1}{3}\right)\right] + \frac{A(1 - \beta')e^{-\frac{t-t_1}{\tau_{\text{fall}}}}}{1 + e^{-\frac{t-t_0}{\tau_{\text{rise}}}}} \cdot \left[\sigma\left(\frac{t-t_1}{3}\right)\right] \tag{2}$$

where F is the model flux of the light curve and A, t_0, t_1, τ_{fall}, τ_{rise}, β' are free parameters that were tuned for each light curve. We set uniform priors for each of the parameters to constrain them to physically realistic values. We then drew sample parameters of the posterior function using Markov Chain Monte

Carlo (MCMC) sampling. We obtained the times of maximum light $t_{\max}$ along with uncertainties $\sigma_{t_{\max}}$ from the posterior predictive light curves.

We similarly obtained the times of first light by fitting the early light curve at times $t < t_{\max}$ with a power law model as defined in equation 1 from Vallely et al. [16] as follows,

$$F = H(t - t_0).C.(t - t_0)^{\beta_1(1+\beta_2(t-t0))} + f_0, \tag{3}$$

where C, t_0, f_0, β_1, and β_2 are free parameters that were tuned for each light curve, t_0 is the time of first light, and H is the Heaviside step function. We set uniform priors to keep the free parameters within physically realistic ranges and then used MCMC sampling to obtain posterior samples of t_0. We used the mean and standard deviation of these samples as the target labels of the time of first light t_0 and corresponding uncertainty σ_{t_0}, respectively. We use both the mean and uncertainties of the target ages in the loss function of the neural network model.

3.2 Probabilistic Recurrent Neural Network

We developed a probabilistic recurrent neural network (RNN) built using Gated Recurrent Units (GRUs, [3]) to map an input multi-passband light curve matrix, $\boldsymbol{X}_{s(t \leq T)}$, for transient s up to time T onto a probabilistic prediction of the transient age,

$$\boldsymbol{Y}_s^w = \boldsymbol{f}(\boldsymbol{X}_{s(t \leq T)}; \boldsymbol{w}), \tag{4}$$

where $\boldsymbol{w}$ are the parameters (i.e. weights and biases) of the network, and $\boldsymbol{Y}_s^w$ has been parameterized as a Normal distribution with a predicted mean age $\hat{y}_s^w$ and a predicted uncertainty $\hat{\sigma}_s^w$ for a particular set of network weights w. We have built two distinct RNN models: one to predict the time of first light with a predicted mean $\hat{y}_s^w = \hat{t}_0^w$ and uncertainty $\hat{\sigma}_s^w = \hat{\sigma}_{t_0}^w$ and another to predict the time of maximum light with a predicted mean $\hat{y}_s^w = \hat{t}_{\max}^w$ and uncertainty $\hat{\sigma}_s^w = \hat{\sigma}_{t_{\max}}^w$. Each model is designed to regress over past data to make real-time predictions that get updated as new data are observed. The model architectures and inputs are identical, with the only difference being the output target variable.

The recurrent neural network architecture comprises two GRU layers, each consisting of 50 units with ReLU activation functions. These GRU layers are unidirectional, meaning only information from past timesteps are used for prediction. This is to ensure that the network makes temporally causal real-time predictions that update as additional data becomes available. The final layer of the network is a time-distributed fully-connected probabilistic layer. It connects all 50 neurons from the second GRU layer to an output vector $\boldsymbol{Y}_s^w$ of size 2 at each timestep for the predictive mean and standard deviation. Each layer in the network includes batch normalization and dropout regularization with a dropout rate of 20%. The input to the network is the matrix $\boldsymbol{X}_s$.

3.2.1 Uncertainty Quantification

Applications of deep learning to astronomy have often avoided uncertainty quantification in favor of point estimation. Here, we exemplify a simple methodology for quantifying uncertainty for any NN regression problem.

To characterize the intrinsic uncertainty of our network's ability to predict age, we build a probabilistic network and parameterize the output as a Normal distribution. Therefore, instead of the typical neural network that only predicts a point estimate, our network outputs a predictive mean $\hat{y}_s^w$ and a standard deviation $\hat{\sigma}_s^w$ for a particular set of weights $\boldsymbol{w}$. The uncertainty $\hat{\sigma}_s^w$ is helpful to model the intrinsic error of our model, such that even if we had no measurement error and an infinite training set, there would still be some discrepancy between our model predictions and the true age of the transient.

We then also quantify the epistemic uncertainty of our model's predictions of $\boldsymbol{Y}_s^w$ by making an approximate Bayesian neural network. We employ an approach by Gal & Ghahramani [6] called Monte Carlo (MC) dropout sampling, which has been shown to provide posterior samples of the network parameters when the commonly used dropout regularization technique is applied during inference instead of just during training. We collect the results of stochastic forward passes through the network as approximate posterior draws of $\boldsymbol{Y}_s^w$ and use the mean and standard deviation of these draws as our marginal predictive mean $\hat{y}_s$ and predictive uncertainty $\hat{\sigma}_s$.

The predictive distribution for the underlying latent age $y_s(\boldsymbol{w})$ of a transient is

$$\mathcal{P}(y_s | \boldsymbol{X}_{s(t \leq T)}, \boldsymbol{w}) = \mathcal{N}\left(y_s(\boldsymbol{w}) \mid \hat{y}_s^w, \hat{\sigma}_s^w\right), \tag{5}$$
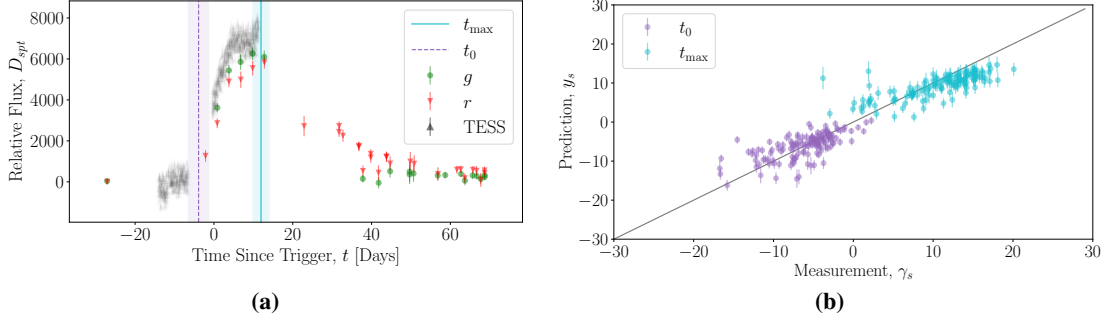
3

Figure 1: (a) An example light curve of a transient (SN2018hyy) in our dataset. The vertical cyan and dashed purple lines represent the predicted time of maximum light $t_{\max}$ and explosion light t_0, respectively, with the shaded bar illustrating the predicted uncertainties σ_s. (b) A comparison of the predicted and measured ages of all transients in the test set given partial light curve data at 20 days after trigger.

where the latent age $y_s(\boldsymbol{w}) = \hat{y}_s^w + \epsilon_s^w$ is modeled by the neural network predictive mean $\hat{y}_s^w$ with an associated error $\epsilon_s^w \sim \mathcal{N}\left(0, \hat{\sigma}_s^w\right)$ that is a zero-mean Gaussian random variable with standard deviation $\hat{\sigma}_s^w$.

Next, a generative model of the observed age γ_s is derived by adding a measurement error to the latent age, $\gamma_s = \hat{y}_s^w + \epsilon_s^w + \zeta_s$, where we assume that the measurement error of the age $\zeta_s \sim \mathcal{N}\left(0, \eta_s\right)$ is a zero-mean Gaussian random variable with standard deviation η_s. Then, we define the cost function of the neural network as the likelihood function,

$$\mathcal{P}(\gamma_s | \boldsymbol{X}_{s(t \leq T)}, \boldsymbol{w}) = \mathcal{N}\left(\gamma_s \mid \hat{y}_s^w, \sqrt{\hat{\sigma}_s^w + \eta_s}\right) \qquad (6)$$

where we include both predicted and measured uncertainties η_s in quadrature. The Bayesian posterior distribution over the weights is

$$\mathcal{P}(\boldsymbol{w}|\boldsymbol{X}) \propto \mathcal{P}(\boldsymbol{w}) \prod_s \prod_T \mathcal{P}(\gamma_s | \boldsymbol{X}_{s(t \leq T)}, \boldsymbol{w}), \qquad (7)$$

where we take the product of the likelihood over all times and all transients in the training set, and the prior is defined as a zero-mean Normal distribution. Finally, to make a prediction of the age of a transient, we evaluate the posterior predictive distribution of the age as follows,

$$\mathcal{P}(y_s | \boldsymbol{X}_{s(t \leq T)}) = \int \mathcal{P}(y_s | \boldsymbol{X}_{s(t \leq T)}, \boldsymbol{w}) \mathcal{P}(\boldsymbol{w}|\boldsymbol{X}) d\boldsymbol{w}, \qquad (8)$$

where we marginalize over the weights of the network by integrating the product of the predictive distribution of the age (first term in the integrand, defined in equation 5) and the posterior distribution over the network weights (second term in the integrand, defined in equation 7). The integral is intractable, so we approximate it using MC dropout [6]. Each inference produces a sample of the posterior; thus, we draw 100 samples i each with a predicted age $y_{s,i}(\boldsymbol{w}_i) \sim \mathcal{N}(\hat{y}_s^w, \hat{\sigma}_s^w)$ by executing 100 forward passes of the neural network for a given input. We compute the marginal predictive mean $\hat{y}_s$ and predictive uncertainty $\hat{\sigma}_s$ as the sample mean and standard deviation of the 100 samples, and report this for each transient.

4 Results and Discussion

We have developed two distinct RNN models, one to predict the time of first light t_0 and the other to predict the time of maximum light $t_{\max}$. We trained the neural network on $\sim 80\%$ of the initial dataset (663 transients), and then validated the performance of the models on the remaining 142 transients ($\sim 20\%$ of the dataset). We drew 100 posterior sample predictions of $\boldsymbol{Y}_s^w$ from our network and plotted the mean and standard deviation of our predictions at 20 days after trigger against the measurements in Fig. 1b. The predictions closely align with the measurements, as evidenced by the scaled errors depicted in Fig. 2a which are distributed near zero. At the time of trigger, the mean absolute error is $\mathrm{MAE}_{t_0} = 2.1$ days and $\mathrm{MAE}_{t_{\max}} = 4.2$ days for the two models. Considering that the average cadence for the ZTF observations is 3 days, our algorithm impressively predicts the correct age (since explosion time) within less than the time between ZTF observations for 68% of transients and within 6 days for 95% of transients.

We examine how our methods perform at early times, and how they improve as new data along the light curves become available. We emphasize that, unlike previous work which predicts the time of maximum
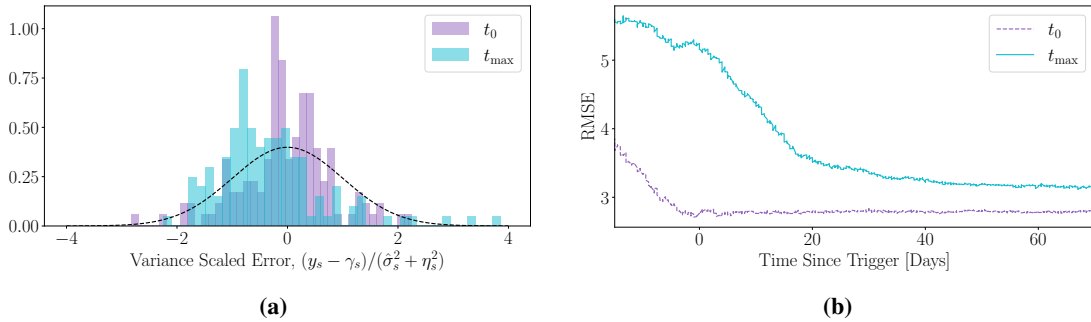
4

Figure 2: (a) The distribution of the error across all time steps scaled by the predictive and measured variances. A unit Gaussian $\mathcal{N}(0,1)$ is shown as the black dashed line to evaluate the effectiveness of the predicted uncertainties. The standard deviations of the scaled errors are 0.99 and 0.81 for the $t_{\max}$ and t_0 models, respectively, and the means are both very close to zero. (b) A plot of the root-mean-square error (RMSE) between the predicted and measured values of the ages. The RNN makes a prediction at every time-step given light curve data up to time T, and we plot the RMSE at each time step. The plot indicates that the predictions improve as more data along the light curves become available.

retrospectively, our work is the first to show results as a function of time and to achieve successful predictions at early times. In Fig. 2b, we plot the root-mean-square error (RMSE) at every time-step outputted by our RNN models. The predictions of the time of first light improve rapidly as more light curve data is recorded by the model, and plateaus at the time of trigger. Thus, as soon as a transient is first detected by ZTF, we can accurately determine the age with an RMSE less than 2.8 days. Predictions of the time of maximum have a slightly larger RMSE, and as expected, improve more after maximum light, which occurs for most transients within 20 days after trigger. For a baseline comparison, we compared our results with the simpler parametric models described in section 3.1. The parametric models had poor performance on partial light curve data and could not compete with the speed of inference of our RNN models.

Overall, this paper presents the first method for identifying the age of a transient in real-time. It is scalable to surveys as large as LSST, with an inference time of 1.7 milliseconds per execution and 0.17 seconds per object for recording 100 posterior samples when run on a single NVIDIA T4 GPU. In future work, we plan to implement this algorithm into the real-time Young Supernova Experiment survey (YSE, [8]). Further work could also investigate utilizing transformers instead of RNNs for the model architecture. We briefly experimented with transformers during this research but found no improvement in accuracy. The limited size of our dataset may be a contributing factor, as a larger dataset could take better advantage of the training speed of transformers. We have evaluated our method on a set of real transient events from the Zwicky Transient Facility (ZTF) and the Transiting Exoplanet Survey Satellite (TESS), and demonstrated the effectiveness of our approach in identifying young transients with high accuracy and quantifiable uncertainties. This work, in combination with real-time transient classifiers, will be critical for discovery and enabling a deeper understanding of progenitor systems in the new era of large-scale astronomical surveys.

Acknowledgments and Disclosure of Funding

References

[1] Allam Tarek J., Peloton J., McEwen J. D., 2023, arXiv e-prints, p. arXiv:2303.08951

[2] Bellm E. C., et al., 2019, Publications of the Astronomical Society of the Pacific, 131, 018002

[3] Cho K., van Merrienboer B., Gulcehre C., Bahdanau D., Bougares F., Schwenk H., Bengio Y., 2014, in Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). Association for Computational Linguistics, pp 1724–1734, doi:10.3115/v1/D14-1179

[4] Foley R. J., Mandel K., 2013, Astrophysical Journal, 778, 167

[5] Gagliano A., Narayan G., Engel A., Carrasco Kind M., LSST Dark Energy Science Collaboration 2021, Astrophysical Journal, 908, 170

[6] Gal Y., Ghahramani Z., 2016, in Proceedings of the 33rd International Conference on International Conference on Machine Learning - Volume 48. ICML'16. JMLR.org, p. 1050–1059

[7] Ivezić Ž., et al., 2019, Astrophysical Journal, 873, 111

[8] Jones D. O., et al., 2021, Astrophysical Journal, 908, 143

[9] Kasen D., 2010, Astrophysical Journal, 708, 1025

[10] Masci F. J., et al., 2023, arXiv e-prints, p. arXiv:2305.16279

[11] Muthukrishna D., Narayan G., Mandel K. S., Biswas R., Hložek R., 2019, Publications of the Astronomical Society of the Pacific, 131, 118002

[12] Muthukrishna D., Mandel K. S., Lochner M., Webb S., Narayan G., 2022, Monthly Notices of the Royal Astronomical Society, 517, 393

[13] Pimentel Ó., Estévez P. A., Förster F., 2023, Astronomical Journal, 165, 18

[14] Ricker G. R., et al., 2015, Journal of Astronomical Telescopes, Instruments, and Systems, 1, 014003

[15] Sánchez-Sáez P., et al., 2021, Astronomical Journal, 161, 141

[16] Vallely P. J., Kochanek C. S., Stanek K. Z., Fausnaugh M., Shappee B. J., 2021, Monthly Notices of the Royal Astronomical Society, 500, 5639

[17] Villar V. A., Cranmer M., Berger E., Contardo G., Ho S., Hosseinzadeh G., Lin J. Y.-Y., 2021, Astrophysical Journal Supplement Series, 255, 24