IMBALANCE-AWARE LEARNING FOR DEEP PHYSICS MODELING

Anonymous authors
Paper under double-blind review

ABSTRACT

In various fields of natural science, there exists a high demand for accurate simulations of physical systems. For example, the weather forecasting requires a large-scale simulation of physical systems described by partial differential equations (PDEs). To reduce the computational cost, recent studies have attempted to build a coarse-grained model of the systems by using deep learning. Many training strategies for deep learning have developed for images or natural languages, but they are not necessarily suited for physical systems. A physical system demonstrates similar phenomena in most points (e.g., sunny days) but exhibits a drastic behavior occasionally (e.g., typhoons). Roughly speaking, a physical system dataset suffers from the class imbalance, whereas previous studies have rarely focused on this aspect. In this paper, we propose an imbalance-aware loss for learning physical systems, which resolves the class imbalance in a physical system dataset by focusing on the hard-to-learn parts. We evaluated the proposed loss using physical systems described by PDEs, namely the Cahn-Hilliard equation and the Korteweg-de Vries (KdV) equation. The experimental results demonstrate that models trained using the proposed loss outperform baseline models with a large margin.

1 INSTRUCTION

Simulations of physical systems are essential in various fields of natural science, such as molecular biology (Polynikis et al., 2009), plasma physics (Tang & Chan, 2005), planetary and space science (Teixeira et al., 1999), and meteorology (Weyn et al., 2021). To reduce the high computational cost of the simulations, recent studies have attempted to build an abstract, coarse-grained, and light-weighted model of the systems by using deep learning (Greydanus et al., 2019; Chen et al., 2018). Since most systems are described by partial differential equations (PDEs) (Stephen, 2000; Kochkov et al., 2021; Weyn et al., 2021), some studies have implemented the prior knowledge about the input space (Raissi et al., 2019) and the underlying geometric structure (Matsubara et al., 2020), thereby improving the modeling and simulation accuracy.

Many training strategies for deep learning have developed for images or natural languages (Devlin et al., 2019), but they are not necessarily suited for physical systems. A physical system demonstrates similar phenomena in most points but exhibits a drastic behavior occasionally. In the case with the weather forecasting, most days are sunny, but occasionally there are typhoons. Hence, the dataset for the weather forecasting suffers from the class imbalance even through no class label is given. Previous studies have rarely focused on this aspect, and existing training strategies for class-imbalanced datasets are inapplicable to learning of physics systems (Felzenszwalb et al., 2010; Shrivastava et al., 2016; Bulo et al., 2017).

In this study, we propose *imbalance-aware loss*, which puts weight on subsets of a given dataset where a model to train cannot predict the future states well. More specifically, we apply the softmax function to the values of the loss function corresponding to subsets, and treat the outputs as the weights for learning the subsets. We evaluated the proposed loss on physical systems described by PDEs, namely the Cahn-Hilliard (CH) equation and the Korteweg-de Vries (KdV) equation. The experimental results demonstrate that models trained using the proposed loss predict the future states much better than baseline and comparison models.

2 BACKGROUND

Class Imbalance. There exists a class imbalance in a dataset if it contains more samples of a class than other classes. With a class-imbalanced dataset, assigning the label of majority class to all samples results in a high accuracy, but such result is almost meaningless. Learning from imbalanced data is one of the key challenges that the machine learning community is facing (He & Garcia, 2009). Typically, a model puts larger weight on or chooses more frequently samples of the minority classes for learning than samples of the majority class.

We face a class imbalance also in object detection, where objects of a class appear more frequently than objects of other classes (Felzenszwalb et al., 2010). Also in this situation, an object detector has a risk of ignoring objects of the minority class. A typical solution is hard negative mining, with which a model put weight on an object that the model cannot detect well (Shrivastava et al., 2016). There exist many sophisticated schemes for sampling and weighting (Bulo et al., 2017).

The focal loss is one of the state-of-the-art weighting strategies for object detection Lin et al. (2017). Let $p_t \in [0, 1]$ denote the probability that a model's estimated class label is correct, and let $\mathcal{L}$ denote the original loss function given p_t. Then, the focal loss is given by $\mathrm{FL}(p_t) = (1 - p_t)^\gamma \mathcal{L}(p_t)$, where $\gamma \geq 0$ is a tunable focusing parameter. The focal loss puts weight on hard negatives (i.e., samples with smaller p_t) and ignores the vast number of easy negatives (i.e., samples with larger p_t).

Imbalanced Data in Physical System. Some physical systems exhibit situations similar to the class imbalance. Figure 1 shows an example taken from a simulation of the CH equation, which is a model of melts of polymers and separation of phase. The polymers form phases that change only gradually at the most portions, while other locations show boundaries, separation, and collision of phases. Therefore, the dataset of the CH equation can be classified into several categories according to the characteristics of the state. Moreover, some categories appear frequently, and some others rarely appear, implying the existence of the class imbalance. Even though Lin et al. (2017) proposed the focal loss to address the class imbalance in object detection, it is inapplicable to time series data including data of physical systems.

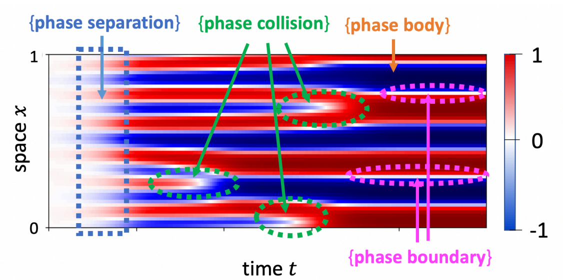

Figure 1: Simulation data of the Cahn-Hilliard equation. There exist several categories according to the characteristics of the state.

3 METHOD

Imbalance-Aware Loss. A physical system follows an ordinary differential equation (ODE) $\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = f(\boldsymbol{u})$, where $\boldsymbol{u}$ is an M-dimensional state. Given a state $\boldsymbol{u}^{t_i}$ at time t_i, the purpose is to predict the successive state $\boldsymbol{u}^{t_{i+1}}$ at time t_i by integrating the ODE. To reduce the computational cost for prediction, many studies have employed deep learning to approximate the time-derivative f or a discrete-time map $\Phi(\boldsymbol{u}) = \int_{t_i}^{t_{i+1}} f(\boldsymbol{u}(s))\mathrm{d}s$ (Greydanus et al., 2019; Raissi et al., 2019). For training, state $\boldsymbol{u}$ is collected randomly form multiple time points in multiple trials; the set of the states is called a minibatch and fed to a model at one time. Let B denote the number of samples in a minibatch, and $\boldsymbol{u}_b$ $(b = 0, \ldots, B-1)$ denote a state in the minibach. Let $\boldsymbol{u}'_b$ denote the successive state of the state $\boldsymbol{u}_b$, which is the target for a model to predict. Let $\hat{\boldsymbol{u}}'_b$ denote the set of states predicted by a model. A typical loss function is the mean squared error (MSE) $\mathcal{L}(\boldsymbol{u}_b) = \frac{1}{MB}||\boldsymbol{u}'_b - \hat{\boldsymbol{u}}'_b||_2^2$.

In this section, we propose the imbalance-aware loss to address the class imbalance in data of physical systems. In a computer simulation, a PDE system is spatially discretized and treated as an ODE. In a discretized PDE system, a state $\boldsymbol{u}_b$ is given as a M-dimensional vector $\boldsymbol{u}_b = (u_{b,0}, u_{b,1}, \cdots, u_{b,M-1})$. Using the element-wise squared error $l_{b,m} = ||u'_{b,m} - \hat{u}'_{b,m}||_2^2$, the MSE is rewritten as $\mathcal{L}(\boldsymbol{u}_b) = \frac{1}{MB}\sum_{b=0}^{B-1}\sum_{m=0}^{M-1} l_{b,m}$. Obviously, the squared error of each element has the same weight $\frac{1}{MB}$. The imbalance-aware loss assigns different weight to different element dynami-

Table 1: Results on the PDE systems using model with the Euler method.

Model	Cahn-Hilliard equation States	Cahn-Hilliard equation Energy	KdV equation States	KdV equation Energy
Original	3.50×10^{-2}	1.96×10^{-2}	2.29×10^{0}	7.14×10^{3}
SANN	2.05×10^{-2}	5.35×10^{-3}	2.04×10^{0}	5.89×10^{3}
Proposed ($T=10^{-1}$)	$\mathbf{7.03\times10^{-3}}$	$\mathbf{3.97\times10^{-5}}$	$\mathbf{1.38\times10^{0}}$	$\mathbf{1.26\times10^{3}}$
Proposed ($T=10^{2}$)	$\mathbf{8.87\times10^{-3}}$	$\mathbf{4.92\times10^{-5}}$	$\mathbf{8.64\times10^{-1}}$	$\mathbf{8.35\times10^{2}}$

The best and second best results are emphasized by bold fonts.

cally. We apply the softmax function to the squared errors $l_{b,m}$ of all elements ($b=0,\dots,B-1$ and $m=0,\dots,M-1$); specifically,

$$w_{b,m}=\frac{\exp(l_{b,m}/T)}{\sum_{b'=0}^{B-1}\sum_{m'=0}^{M-1}\exp(l_{b',m'}/T)}, \tag{1}$$

where T is a temperature parameter. Using the weight $w_{b,m}$, we propose the imbalance-aware loss

$$\mathrm{IAL}(\boldsymbol{u}_b)=\textstyle\sum_{b=0}^{B-1}\sum_{m=0}^{M-1}w_{b,m}l_{b,m} \tag{2}$$

The weight $w_{b,m}$ becomes more even as T increases, while the weight $w_{b,m}$ becomes emphasized more as T decrease. With the limit $T\to 0$, we obtain $\mathrm{IAL}(\boldsymbol{u}_b)=\max_{b,m}(l_{b,m})$. With the limit $T\to\infty$, the imbalance-aware loss IAL converges to the ordinary loss $\mathcal{L}$.

What Does Imbalanced-Aware Loss Emphasize? We show which parts of the CH equation data in Figure 1 are emphasized by the imbalance-aware loss. Figure 2 shows the MSE between the ground truth $\boldsymbol{u}'_b$ and the states $\hat{\boldsymbol{u}}'_b$ predicted by one step using a model trained with a limited number of iterations (see Section 4 for details of the experiment). The MSE is multiplied by 10^5 for visibility. A darker colored pixel implies a large error. We find that larger errors appear at areas categorized into phase separation, collision, and boundary, indicating that these areas are difficult to learn. The imbalance-aware loss can prevent the bias of the learning process to easy-to-learn areas (i.e., phase bodies) because it gives larger weights to the difficult areas, thereby handling the class imbalance.

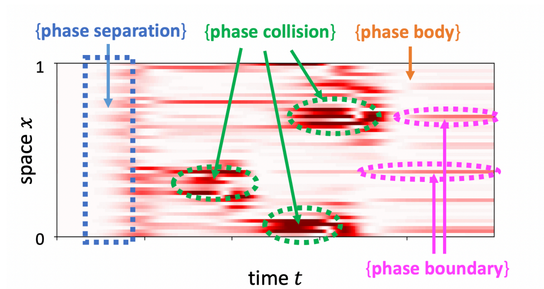

Figure 2: The squared error between ground truth and per-step estimates for the Cahn-Hilliard equation data used in Figure 1 ($\times10^5$). There are four main categories according to the characteristics of the state.

4 Experiments and Results

Experimental Setting. We evaluated our method on PDE systems, namely the CH equation and the KdV equation, which were examined by Matsubara et al. (2020). We used the HNN++ for the CH system and the Neural ODE (NODE) for the KdV system; NODE treats the output of a neural network as the time derivative of the input (Chen et al., 2018), where as the HNN++ takes the gradient of a neural network for obtaining the time derivative and thereby guarantees the underlying geometric properties. To emphasize the effectiveness of the imbalance-aware loss, we employed simpler integration methods, namely the Euler method and the explicit midpoint method (RK2). For comparison, we also trained these models using the ordinary loss function and the loss function proposed by the SANN Liang et al. (2022), which resamples stiff parts. The details of the experiments are summarized in Appendix A.

Results. After training, we solved the initial value problems and obtained the MSE between the ground truth and states predicted by trained models. In addition, we also obtained the MSE in the system energy. Table 1 summarizes the results, where we provided a median of 15 trials. The imbalance-aware loss improved the prediction accuracy by a large margin for both the CH equation and the KdV equation. We found that the prediction by a model trained using SANN is more accurate than that of ordinary loss and still worse than that of the imbalance-aware loss. This fact indicates that focusing on stiff portions is important but not sufficient. SANN can put weights on collisions of phases, but cannot do it on phase boundaries. Even when using RK2, the imbalance-aware loss achieved the best results (see Appendix for details). We also visualized the prediction results of the CH equation and the KdV equation with the Euler method, as depicted in Figures 3 and 4, respectively. Without the imbalance-aware loss, the prediction exhibits larger errors and awkward noises for both the CH equation and the KdV equation.

Figure 5 shows the results when varying the temperature T. The best result was obtained with $T = 10^{-1}$ for the CH equation and with $T = 10^{2}$ for the KdV equation. While the optimal value depends on the datasets and integrators, the imbalance-aware loss significantly outperformed the comparison methods over a wide range of the temperature T, namely between 0 and 10^2.

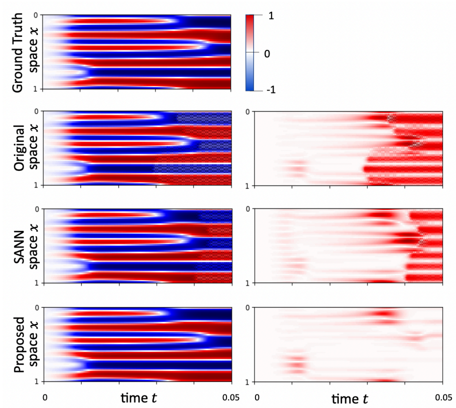

Figure 3: The CH equation predicted by the HNN++ with the Euler method. (left) Ground truth $\boldsymbol{u}$ and predicted state $\hat{\boldsymbol{u}}$. (right) Error. The results of the imbalance-aware loss is with $T = 0$.

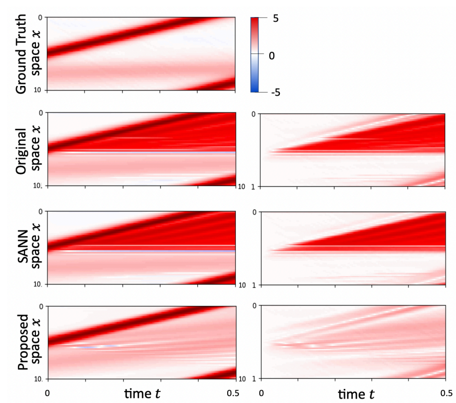

Figure 4: The KdV equation predicted by the NODE with the Euler method. (left) Ground truth $\boldsymbol{u}$ and predicted state $\hat{\boldsymbol{u}}$. (right) Error. The results of the imbalance-aware loss is with $T = 10^2$.

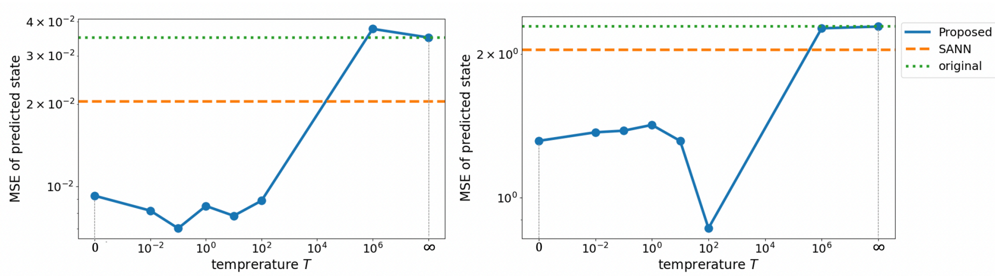

Figure 5: The MSE of (left) the CH equation predicted by the HNN++ with the Euler and (right) the KdV equation predicted by the NODE with the Euler method with varying the temperature T. Both axes are in logarithmic scales.

5 Conclusion

We proposed the imbalance-aware loss as a learning method that addresses the imbalance in time series dataset. We evaluated it on the CH equation and the KdV equation, and demonstrated that a model trained using the proposed loss outperformed the baselines by a large margin.

References

S. R. Bulo, G. Neuhold, and P. Kontschieder. Loss Max-Pooling for Semantic Image Segmentation. In *Conference on Computer Vision and Pattern Recognition*, pp. 1–14, 2017.

T. Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud, R. T. Q. Chen, Y. Rubanova, J. Bettencourt, and D. Duvenaud. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, pp. 1–19, 2018.

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *NAACL-HLT*, pp. 4171–4186, 2019.

P. F. Felzenszwalb, R. B. Girshick, and D. McAllester. Cascade Object Detection with Deformable Part Models. In *Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2010.

S. Greydanus, M. Dzamba, and J. Yosinski. Hamiltonian Neural Networks. In *Advances in Neural Information Processing Systems*, pp. 1–16, 2019.

H. He and E. A. Garcia. Learning from Imbalanced Data. In *IEEE Transactions on Knowledge and Data Engineering*, pp. 1263–1284, 2009.

D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, pp. 1–15, 2015.

D. Kochkov, J. A. Smith, A. Alieva, Q. Wang, M. P. Brenner, and S. Hoyer. Machine learning accelerated computational fluid dynamics. In *Proceedings of the National Academy of Sciences of the United States of America*, pp. 1–13, 2021.

S. Liang, Z. Huang, and H. Zhang. Stiffness-Aware Neural Network for Learning Hamiltonian System. In *The International Conference on Learning Representations workshop*, 2022.

T.Y. Lin, P. Goyal, R. Girshick, K. He, and P. Doll'ar. Focal loss for dense object detection. In *International Conference on Computer Vision*, volume 42, pp. 318–327, 2017.

T. Matsubara, A. Ishikawa, and T. Yaguchi. Deep Energy-Based Modeling of Discrete-Time Physics. In *Advances in Neural Information Processing Systems*, pp. 1–24, 2020.

A. Polynikis, S.J. Hogan, and M. di Bernardo. Comparing different ODE modelling approaches for gene regulatory networks. *Journal of Theoretical Biology*, pp. 511–539, 2009.

M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. In *Journal of Computational Physics*, pp. 686–707, 2019.

A. Shrivastava, A. Gupta, and R. Girshick. Training Region-based Object Detectors with Online Hard Example Mining. In *Conference on Computer Vision and Pattern Recognition*, pp. 1–9, 2016.

B. P. Stephen. *Turbulent Flows*. Cambridge university press, 2000.

W.M. Tang and V.S. Chan. Advances and challenges in computational plasma science. *Plasma Physics and Controlled Fusion*, 47:R1–R34, 2005.

J. Teixeira, J.W. Lockhart, M.W. Liemohn, and X. Jia. Stable Schemes for Partial Differential Equations. *Journal of Computational Physics*, pp. 403–417, 1999.

J. A. Weyn, D. R. Durran, R. Caruana, and N. Cresswell-Clay. Sub-seasonal forecasting with a large ensemble of deep-learning weather prediction models. *Journal of Advances in Modeling Earth Systems*, 13:1–23, 2021.