SF2A 2021
A. Siebert, K. Baillié, E. Lagadec, N. Lagarde, J. Malzac, J.-B. Marquette, M. N'Diaye, J. Richard, O. Venot (eds)

CLASSIFICATION OF COSMOLOGICAL MODELS FROM THE INTERNAL PROPERTIES OF DM HALOS BY USING MACHINE LEARNING

R. Koskas[1] and J. M. Alimi[1]

Abstract. We are interested in detecting the cosmological imprint on properties of present dark matter halos by using Machine Learning methods. We analyse the halos formed in Dark Energy Universe Simulations using several dark energy models (ΛCDM, Quintessence Ratra Peebles), whose parameters were chosen in agreement with both CMB and SN Ia data. Their resulting halos are thus extremely close from one cosmological model to another. However, we have shown that machine learning techniques can be implemented to determine the cosmological model in which each halo was formed: we associate to each present day halos from ΛCDM and RP CDM ellipsoidal mass and shape profiles, defined to efficiently keep track of the matter distribution anisotropies and, then, we experimentally show that those quantities allow a properly trained learning device to find the dark energy model of the Universe within which these halos have grown. Training our device on 40,000 halos of 10^{13} and 10^{14} solar masses, we can correctly classify more than 70% of the halos in the test set. We also study the misleading ML methodological biases, "Clever Hans effects", and the way to fix them.

Keywords: dark energy, machine learning, decision tree, numerical simulations, dark matter halos

1 Introduction

Dark Energy Universe Simulations (Alimi et al. 2010; Rasera et al. 2010; Alimi et al. 2012; Reverdy et al. 2015) is a set of high performance N-body cosmological simulations. From one simulation to another, several dark energy models are assumed, e.g. the dynamical Ratra Peebles (RP) CDM model and the fiducial ΛCDM. We will use in this proceedings the 648 $h^{-1}Mpc$ simulation containing $N = 2048^3$ particles. For each cosmological model, all cosmological parameters are chosen to form a CMB (Spergel et al. 2007) and SN Ia (Kowalski et al. 2008) compatible n-uplet. As a consequence, we study only **realistic** models (Alimi et al. 2010), which are extremely close one to each other - any halo of the ΛCDM Universe strongly overlap its $RPCDM$ counterpart. Therefore, natural questions emerge: **is there any cosmological imprint in the difference between the halos of the two Universes?** In other words, **does the internal structure of halos embed cosmological information ?** Our goal is to extract information about the cosmology (dark energy model) from the matter distribution and the dynamical state of the simulated halos. Whereas a conventional objective would consist in exhibiting a mean behaviour, *i.e.*a quantity whose the average value on a large population simulated halos change significantly (more than 1σ) with the dark energy model, our objective is here more predictive: we want to infer the cosmology from only the internal properties of each individual halo. To be more specific, **each halo is described by a common set of chosen quantitative attributes; we then aim at training an AI to associate to each set of attributes, the dark energy model of the Universe in which the corresponding halo has grown. This is a classification task**. Furthermore, by changing the set of chosen attributes, we will select those that are the most significant from a cosmological point of view.

2 Halo properties computation

For a correct description of a halo, it is necessary to capture the spatial distribution of the matter in it, that is to say its profile. Because DM halos are triaxial ellipsoids rather than isotropic (Despali et al. 2016), a

[1] Laboratoire Univers et Théories, CNRS UMR 8102, Observatoire de Paris, PSL University

thorough profile characterisation requires local density measures - The parameter free Delaunay Tessellation Field Estimator (Cautun & van de Weygaert 2011) provides the local density δ at each particle locus. That allows to remove subhalos. For each density δ_a in a sequence of pre-chosen "points of measure" $(\delta_a)_a$, we consider the δ_a isodensity shell $\mathcal{S}_a = \{k \mid |\frac{\delta(\mathbf{x}^k)}{\delta_a} - 1| \leq 0.1\}$ and we observe (Jing & Suto 2002) that it is approximately an ellipsoid $(\mathcal{E}_a)$ clearly non spherical. Its parameters are computed through the diagonalization the mass tensor of the particles forming the shell $\mathcal{S}_a$: $(M_{ij}^{\mathcal{S}_a})_{1\leq i\,,\,j\leq 3} = \langle x_i x_j \rangle_{\mathcal{S}_a} - \langle x_i \rangle_{\mathcal{S}_a} \langle x_j \rangle_{\mathcal{S}_a}$. Finally, in each fitted ellipsoid $\mathcal{E}_a$, one computes the quantities that will describe halo's structure and its dynamics (the set of attributes) namely the enclosed mass $\mathcal{M}_a$ (which is thus a multiple of the particle mass m_p), the length of $\mathcal{E}_a$ axis, the velocity dispersion σ_a^V,... and so on (Koskas & Alimi 2021).

3 Machine learning

Our objective is now to train an AI to associate to a sequence $(\mathcal{M}_a, \mathcal{E}_a, \sigma_a^V, ...)_a$ the dark energy model (fiducial or RP) of the Universe in which the corresponding halo was formed. We use for this propose an ensemble of decision trees , aggregated by gradient boosting (Friedman 2002). However, the use of Machine Learning algorithms induces specific subtle spurious effects. As, we use N-body simulations where all the particles have the same elementary mass m_p and because we have chosen realistic models, Ω_m and thus m_p are different from one dark energy model to another. Now, the $\mathcal{M}_a$'s belong to $m_p^{RP}\mathbb{N}$ or $m_p^{\Lambda}\mathbb{N}$, which do not intersect. So, if the machine detects that all first-cosmology halo masses in the train set are multiple of the same elementary mass (that the machine should determine) and that all second-cosmology halo masses are multiple of another base mass, then the machine will also be able to classify the halos of the test set (simply by looking if their masses are multiple of m_p^{Λ} or m_p^{RP}). In other words, data contain cosmological information of purely arithmetical nature, which would not be reproduced in a real Universe (continuous fields). It is a typical Clever Hans effect.**This kind of effects has been carefully hunted and corrected in order to obtain physically reliable results.** See (Koskas & Alimi 2021) for the way we altered the train set to avoid it.

4 Results and conclusions

Once trained, our AI can be tested individually on each halo in the test set and determines the cosmological model with a 71% probability of success. This result is essentially achieved by using ellipsoidal profile approach rather than the spherical one, and only with mass $(\mathcal{M}_a)$, and shape attributes $(\mathcal{E}_a)$. Those are therefore the most cosmologically impregnated properties of the halos. In Koskas & Alimi (2021) we further discuss the required precision on the attributes for the AI to be predictive. We also explain why velocity dispersion measurements are not sufficient to classify the halos.
As a conclusion, it is possible to read in the halo structure the dark energy model. To do so, one has to finely describe the mass profile through local density computation and **ellipsoidal** approximation for iso-densities. Supposing isotropy when determining the mass profile considerably lowers the result. Also, it is crucial to understand **how** the resulting engine works. In particular, one has to check that the classification is achieved only by **physical** means, ignoring any cosmological clue coming from the numerical nature of the simulation, so that the engine would work for a real Universe.

References

Alimi, J.-M., Bouillot, V., Rasera, Y., et al. 2012, IEEE Computer Soc. Press, CA, USA, SC2012, Article No 73

Alimi, J.-M., Füzfa, A., Boucher, V., et al. 2010, MNRAS, 401, 775

Cautun, M. C. & van de Weygaert, R. 2011, Astrophys Source Code Library

Despali, G., Giocoli, C., & et al, M. B. 2016, MNRAS, 466, 181

Friedman, J. H. 2002, Comput. Stat. Data Anal., 38, 367

Jing, Y. P. & Suto, Y. 2002, Astrophys. J., 574, 538

Koskas, R. & Alimi, J.-M. 2021, in preparation

Kowalski, M., Rubin, D., Aldering, G., et al. 2008, Astrophysical J, 686, 749

Rasera, Y., Alimi, J.-M., Courtin, J., et al. 2010, in AIP proc. 'Invisible Universe International Conference'

Reverdy, V., Alimi, J.-M., Bouillot, V., et al. 2015, IJHPCA, 29, 249

Spergel, D. N., Bean, R., Doré, O., et al. 2007, ApJS, 170, 377