My name is 林學謙, Today, I will talk about Sparse Identification of Nonlinear Dynamics (SINDy).

SINDy is the abbreviation for Sparse Identification of Nonlinear Dynamics.

It is an algorithm that helps to discover dynamics system from measurement data, because in many physical systems, we cannot comprehend all the exact differential equations to describe the system.

So I take the picture from the video, to illustrate how SINDy works,

We have three variables X Y and Z which describe the evolution of this chaotic system, Lorenz system. In addition, there are three constants, sigma, Rho and beta.

Now, we can run the simulations of the lorent system, and collect measurement data from the simulations. We can build up two matrix. The former matrix is for the different variable combination. The last matrix is considered as the coefficient of the all variable.

And what SINDy does is try to apply some kind of a sparse regression, for example LASSO algorithm, elastic net, to find the linear combination of these non-linearity terms in fewest features, representing X dot Y dot and Z dot,

so when we put these results together, we can figure out this nonlinear model with these coefficients, of which terms in the nonlinear features are active, and the identified system is virtually the same with the equations, it has the same structures and the coefficient Sigma ,Rho and Beta.

Finally, it can be the dynamics system model, and we can use it to simulate the future probable behaviour.

In summary, there are some pros and cons from my view.

The first advantage of SINDy is that it is useful for the dynamic system with difficult differential equations.

The second one is sparsity make it identify the most important features in the systems, so it can also reduce the overfitting. Because sparsity means that we have to choose the small subset variables to represent the dynamic system.

Compare to the advantages, the disadvantages of SINDy is easy to observe by its operation.

The theta matrix, we have to predict our possible theta matrix variable in advance, or we cannot get accurate dynamic model.

What’s more, The data quality is take a important role for the SINDy or common algorithm with machine learning. So how to reduce the noisy and discard inappropriate data is important for any data science engineering.

Thank you for your attention.