Supplementary Material for A Distillation-based Future-aware Graph Neural Network for Stock Trend Prediction

I. APPENDICES

Studies have shown that as coefficient of determination R^2 increases, the regression model's prediction accuracy improves[1], where R^2 is the square of the correlation coefficient ρ. Given two random variable X and Y, where X and Y represent stock historical and future information, respectively. The non-stationarity of stock series data causes distribution shifts between X and Y, i.e., $P(X) \neq P(Y)$, resulting in sub-optimal performance of stock predictors that solely consider historical stock information for prediction.

Proof. Given X and Y, $\exists Z = aX + bY$, $\rho^2_{ZY} > \rho^2_{XY}$, where $\rho^2_{ZY} = \frac{Cov(Z,Y)}{\sqrt{Var(Z)}\sqrt{Var(Y)}}$, $\rho^2_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}$, $\sqrt{Var(X)} > 0$, $\sqrt{Var(Y)} > 0$, $\sqrt{Var(Z)} > 0$:

$$\rho^2_{ZY} > \rho^2_{XY} \equiv \left(\frac{Cov(Z,Y)}{\sqrt{Var(Z)}\sqrt{Var(Y)}}\right)^2 > \left(\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}\right)^2 \tag{1}$$

Substitute $Z = aX + bY$ into Eq.(1):

$$\begin{aligned} Eq.(1) &= \left(\frac{Cov(aX+bY,Y)}{\sqrt{Var(aX+bY)}\sqrt{Var(Y)}}\right)^2 > \left(\frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}\right)^2 \\ &= \left(\frac{Cov(aX+bY,Y)}{\sqrt{Var(aX+bY)}}\right)^2 > \left(\frac{Cov(X,Y)}{\sqrt{Var(X)}}\right)^2 \\ &= \left(\frac{aCov(X,Y)+bVar(Y)}{\sqrt{a^2Var(X)+b^2Var(Y)+2abCov(X,Y)}}\right)^2 > \left(\frac{Cov(X,Y)}{\sqrt{Var(X)}}\right)^2 \\ &= \frac{a^2Cov^2(X,Y)+b^2Var^2(Y)+2abCov(X,Y)Var(Y)}{a^2Var(X)+b^2Var(Y)+2abCov(X,Y)} > \frac{Cov^2(X,Y)}{Var(X)} \end{aligned} \tag{2}$$

For convenience, we denote $Var(X)$ as $\mathbf{A}$, $Var(Y)$ as $\mathbf{B}$ and $Cov(X,Y)$ as $\mathbf{C}$, Thus, Equation (2) can be written as:

$$\begin{aligned} Eq.(2) &= \frac{a^2\mathbf{C}^2+b^2\mathbf{B}^2+2ab\mathbf{BC}}{a^2\mathbf{A}+b^2\mathbf{B}+2ab\mathbf{C}} > \frac{\mathbf{C}^2}{\mathbf{A}} \\ &= \mathbf{A}(a^2\mathbf{C}^2+b^2\mathbf{B}^2+2ab\mathbf{BC}) > \mathbf{C}^2(a^2\mathbf{A}+b^2\mathbf{B}+2ab\mathbf{C}) \\ &= b^2\mathbf{B}(\mathbf{AB}-\mathbf{C}^2) > 2ab\mathbf{C}(\mathbf{C}^2-\mathbf{AB}) \end{aligned} \tag{3}$$

$$\begin{aligned} \mathbf{AB}-\mathbf{C}^2 &= Var(X)Var(Y) - Cov^2(X,Y) \\ &= Var(X)Var(Y) - \rho^2_{X,Y}Var(X)Var(Y) \\ &= (1-\rho^2_{X,Y})Var(X)Var(Y) \end{aligned} \tag{4}$$

According to Eq.(4), $\mathbf{AB}-\mathbf{C}^2 > 0$. Eq.(3) can be futrher written as, $b^2\mathbf{B}+2ab\mathbf{C} > 0$. Thus, $\exists a > 0, b > 0$, such that satisfy the inequality. Without loss of generality, $\exists Z = f(X,Y), \rho^2_{ZY} > \rho^2_{XY}$. Therefore, we can conclude from the proof that considering future information in addition to historical information enhances model's predictive performance.

REFERENCES

[1] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation," *Peerj computer science*, vol. 7, p. e623, 2021.