

นิยามของ Concept Drift ใน AI & Machine Learning

ในบทความนี้เราจะแบ่งเนื้อหาออกเป็น 2 หัวข้อได้แก่ 1) นิยามของ Concept Drift 2) ที่มา (ทางคณิตศาสตร์) ของ Concept Drift

นิยามของ Concept Drift

เพื่อความถูกต้องและความเป็นสากลของนิยาม เราขอพูดถึงเนื้อหาในส่วนนี้ด้วยภาษาอังกฤษและใช้สัญลักษณ์ทางคณิตศาสตร์ในการอธิบาย

Given a time period $[0, t]$

A set of sample, denoted as $S_{0,t} = \{d_0, \dots, d_t\}$

Where $d_i = (X_i, y_i)$ is one observation (or a data instance)

X_i is the feature vector

y_i is the label

$S_{0,t}$ follows a certain distribution $F_{0,t}(X, y)$

Concept drift occurs at timestamp $t + 1$,

if $F_{0,t}(X, y) \neq F_{t+1,\infty}(X, y)$, denoted as $\exists t: P_t(X, y) \neq P_{t+1}(X, y)$

ที่มา (ทางคณิตศาสตร์) ของ Concept Drift

จาก $P_t(X, y)$ ที่ได้มาข้างต้น เราสามารถแยกส่วนประกอบได้เป็น 2 ส่วนดังนี้

$P_t(X, y) = P_t(X) \times P_t(y|X)$ ซึ่งพิจารณาได้ว่า Concept Drift สามารถเกิดขึ้นจาก 3 ที่มาดังนี้

ที่มา I: $P_t(X) \neq P_{t+1}(X)$ ในขณะที่ $P_t(y|X) = P_{t+1}(y|X)$ นั่นคือ new data ที่เข้ามา อยู่ใน X บริเวณที่เราไม่ได้เห็นมาก่อนใน historical data แต่ยังสามารถใช้ decision boundary เดิมที่เคยหาไว้จาก historical data ได้

ที่มา II: $P_t(y|X) \neq P_{t+1}(y|X)$ ในขณะที่ $P_t(X) = P_{t+1}(X)$ นั่นคือ new data ที่เข้ามา อยู่ใน X บริเวณที่เราเคยเห็นมาก่อนใน historical data แต่พฤติกรรมของ y กลับแตกต่างจากเดิม การเปลี่ยนแปลงนี้ทำให้ decision boundary ของ new data แตกต่างจากของ historical data และนำไปสู่ประสิทธิภาพที่ลดลงของโมเดล (เราจำเป็นต้องปรับโมเดลตาม)

ที่มา III: $P_t(X) \neq P_{t+1}(X)$ และ $P_t(y|X) \neq P_{t+1}(y|X)$ นั่นคือ new data ที่เข้ามา อาจอยู่/ไม่อยู่ใน X บริเวณที่เราเคยเห็นมาก่อนใน historical data และพฤติกรรมของ y ก็แตกต่างจากเดิม การเปลี่ยนแปลงนี้ทำให้ decision boundary ของ new data แตกต่างจากของ historical data (โดยที่ต้องใช้ความปรานีในการ

อัลกอริทึมเดอลมากกว่า ที่มา II เนื่องจากเราต้องคำนึงถึง X ในบริเวณที่ไม่เคยเจอมาก่อนใน historical data ด้วย)

.

อ้างอิงจากงานวิจัย: Learning under Concept Drift - A Review

.

ต่อไป เราจะพูดถึงประเภทของ Concept Drift กัน (ตอนสุดท้ายก่อนเริ่มอธิบาย technical ของ ส่วนประกอบต่าง ๆ ของ Concept Drift)

.

แล้วพบกันใหม่ สวัสดีครับ

.

ติดต่องาน inhouse training และ AI consult ได้ที่ inbox เพจ @TautologyThailand