UNIVERSITY OF
SOUTH CAROLINA

جامعة ساوث كارولينا بالولايات المتحدة الأمريكية

إطار النمذجة المكدسة

رسالة مقدمة من

المهندس /كريم عز الدين ربيع

كجزء من متطلبات الحصول على

درجة الدكتوراة في تخصص علوم الحاسب

تحت إشراف

Gabriel Terejanu, Major Professor

لجنة تحكيم

John Rose, Committee Member

Marco Valtorta, Committee Member

Jianjun Hu, Committee Member

Andreas Heyden, Committee Member

Cheryl L. Addy, Vice Provost and Dean of the Graduate School

Stacked Modelling Framework

إطار النمذجة المكدسة

تقوم الأطروحة بتقديم إطار عام لعمل نمذجة تنبؤية محتملة استنادًا إلى عمليات مكدسة (Stacked Gaussian) لعمليات Gaussian مدربة بشكل مستقل ، تسمى StackedGP ، تم اقتراحها كإطار عمل للنمذجة في مجال تدريب الآلة (Machine learning) لمواجهة تحديين: النمذجة القائمة على مدخلات المشكلة المراد حلها في ظل عدم اليقين الهيكلي (structural uncertainty) وانتشار عدم اليقين (propagation of uncertainties) بين أجزاء النموذج وصولا للمخرجات المتوقعة من النموذج. ينشأ أحد التحديات الحالية في النمذجة القائمة على المكونات من الطبيعة البارامترية (parametric nature) المجهزة للنماذج الفرعية (sub-models) مع عدم وجود معلومات في معظم الأحيان عن حجم عدم اليقين في البيانات المتاحة. على الرغم من استخدام طرق متعددة لأخذ عينات (sampling methods) لنشر عدم اليقين البارامترى (propagate the parametric uncertainty) للمخرجات المتوقعة من النموذج، إلا أن عدم اليقين في تنبؤات النموذج قد لا يزال يتم التقليل من شأنها عن طريق تجاهل عدم اليقين من نموذج. يعتمد النموذج المقترح على شبكة من عمليات غاوس المدربة بشكل مستقل مصحوبة بطريقة رياضية تقريبي للحصول على للمخرجات المتوقعة من النموذج الذي يعتمد على عدة نماذج الفرعية (sub-models). يوفر النموذج المقترح تقديرًا لمخرجات النموذج مع عدم اليقين الكمي. هذا يعزز القدرة التحليلية لعملية غاوسين ذات مدخلات غير مؤكدة عند استخدام kernals أسية مربعة ومتعددة الحدود. يمكن أن يمتد شبكة StackedGP إلى أي عدد من العقد والطبقات وليس له أي قيود في اختيار kernel مناسب لعقد الإدخال. تُظهر النتائج العددية فائدة استخدام StackedGP للتعلم من مجموعات البيانات المتعددة ونشر عدم اليقين في الكميات ذات الأهمية. يأتي هذا مع تعقيد حسابي أقل مقارنة بالطرق متعددة المخرجات.

تم تقديم إطار عملية Gaussian المكدسة التحليلية (StackedGP) إلى واحدة من المشاكل البيئية الصعبة ، وتوقع تركيزات الأفلاتوكسين مع عدم اليقين الكمي. تُستخدم عمليات غاوس العشوائية لنمذجة مدخلات درجة الحرارة والنشاط المائي لدفع عملية غاوس للأفلاتوكسين. يتم استخدام مخطط تحليلي لحساب توقعات تنبؤات الأفلاتوكسين عن طريق نشر عدم اليقين من خلال عملية Gaussian المكدسة. تُستخدم بيانات المجال التاريخية من ثلاث حالات مختلفة للتحقق من صحة تنبؤات النموذج على مستوى مكون عملية Gaussian وعلى مستوى النظام لعملية Gaussian المكدسة بأكملها. تم اقتراح اشتقاق جديد لتوسيع الإطار التحليلي عن طريق حساب التغاير في StackedGP بين المواقع الجغرافية المختلفة. يساعد هذا في تحديث التنبؤات السابقة لـ StackedGP باستخدام القياسات الميدانية ، ويوفر الفرصة لربط StackedGP بتقنيات تحسين Bayesian لتحديد المناطق التي يمكن فيها الحصول على قياسات ميدانية جديدة. وبالتالي ، يمكن توسيع StackedGP المقترح بسهولة دمج البيانات البيئية الأخرى مثل التقدُّم والرطوبة. يمكن أيضًا توسيع إطار العمل التنبئي المقترح لمعالجة المراقبة في الوقت الحقيقي لمجموعة واسعة من السموم الفطرية في المحاصيل. يمكن استخدام المعلومات الناتجة عن الأفلاتوكسين StackedGP لتصميم إدارة محاصيل مستدامة عن طريق تقليل كمية المواد الكيميائية المستخدمة في زراعة المحاصيل.

في الجزء الثالث من الأطروحة ، نقدم أيضًا إطار عمل GPالمكدس للتطبيق الكيميائي. أولاً ، قمنا بمقارنة النماذج الخطية مقابل نماذج التعلم الآلي المتقدمة للتنبؤ بطاقات الحالة الانتقالية (TS) باستخدام أكثر من ١٣٣٠ مجموعة واصف من خلال النظر في قاعدة بيانات لطاقات الحالة الانتقالية عبر الأسطح المعدنية (Cu، Rh ، Ru ، Pd ، Pt ،Ni) لنزع الكربوكسيل و نزع الكربوكسيل من حمض البروبيونيك. يظهر التحليل أن النظر في عدد الروابط والواصفات المعدنية يمكن أن يساعد في تحقيق MAE أقل لكل من النماذج الخطية وغير الخطية. إلى جانب ذلك ، لا يمكن للنماذج المعقدة غير الخطية تحقيق نتائج ذات دلالة إحصائية أفضل من أفضل النماذج الخطية. بالإضافة إلى ذلك ، نناقش العديد من مناهج تجميع التفاعل الأولية التي توضح أن نهج النموذج الواحد يمكن أن يؤدي بشكل مشابه لنهج الخطوة الواحدة والذي قد يكون مفيدًا عندما تكون بيانات التفاعل الخاصة بمعادن مختلفة مفقودة. تم التوصل إلى نفس النتيجة في دراسة البيانات المفقودة التي أظهرت أن الواصفات التقليدية (المنتج أو الطاقة المتفاعلة) ستعطي MAE أعلى مقارنة بتلك التي تستخدم أيضًا معلومات تعداد الروابط والواصفات المعدنية. يظهر هذا في نماذج Ridge / GP_RBF للواصفات Er و ١Ep و ٢Ep و Br و ٢Bp و ECH3CH التي تعمل بشكل جيد للغاية. في هذا العمل أيضًا ، أظهرنا أنه يمكن استخدام GP المكدس لبناء خط أنابيب للتنبؤ بطاقات الحالة الانتقالية بأداء عالٍ جدًا مقارنة بحسابات DFT ونماذج التلال المكدسة. يمكن تمديد هذا العمل من خلال دراسة المزيد من مجموعات الواصفات للتنبؤ بطاقات الحالة الانتقالية. أيضًا ، بناء قطع البركان لدراسة TOF هو الخطوة التالية لهذا العمل. بالإضافة إلى ذلك ، يمكن تمديد هذا العمل للأسطح المعدنية الأخرى.