Information-theoretic partition-based models for interpretable machine learning

Yang, L.

Samenvatting

Overal waar digitale systemen zijn worden gegevens vastgelegd. Websites en mobiele applicaties registreren bijvoorbeeld hoe mensen die gebruiken. Sensoren meten en registreren parameters (zoals temperatuur) van het productieproces van industriële producten. Financiële softwaresystemen registreren transacties van financiële activiteiten. Op dezelfde manier registreren zorginformatiebeheersystemen de condities van patiënten.

Dergelijke gegevens kunnen worden gebruikt om informatie over het onderliggende fysieke proces dat deze data 'genereert' bloot te leggen. Onderzoek op het gebied van *data mining* en *machine learning* houdt zich bezig met het ontwikkelen van modellen en algoritmen die effectief en efficiënt informatie uit een dataset kunnen extraheren.

Zo kan informatie die wordt gevonden in de medische dossiers van grote aantallen patiënten in een ziekenhuis gebruikt worden om te begrijpen waarom een bepaalde conditie zeer risicovol is voor één type patiënt maar niet voor de rest. Daarnaast kunnen monitorsystemen ontwikkeld worden om medisch personeel te waarschuwen voor gevaarlijke situaties bij gehospitaliseerde patiënten.

Deze dissertatie richt zich op het ontwikkelen van nieuwe methoden die *partitie-gebaseerde modellen* uit gegevens kunnen construeren. Door een dataset te partitioneren in subgroepen, waar elke subgroep homogeen is vanuit bepaalde perspectieven, extraheert een op partities gebaseerd model patronen in de data, zoals patronen die aangeven welke typen patiënten in ziekenhuizen risico lopen op bepaalde ziektes. Het maakt ook data-gedreven voorspellingen mogelijk, zoals het activeren van alarmen voor potentieel gevaarlijke situaties. Het doel is dus om data mining en machine learning methoden te ontwikkelen die de data 'juist' partitioneren—het concept van *juistheid* in deze dissertatie is gedefinieerd op basis

van een informatie-theoretische benadering.

Specifiek richten we ons op het bestuderen van partitie-gebaseerde modellen voor verschillende taken, inclusief interpreteerbare classificatie met meerdere klassen, discretisatie en samenvatten van data, en het leren van afhankelijkheidsstructuren. Voor interpreteerbare classificatie met meerdere klassen ontwikkelen we een interpreteerbaar machine learning algoritme dat *probabilistische regels* uit gegevens kan extraheren. Een probabilistische regel zou bijvoorbeeld kunnen zijn: *Als het weer mistig is en de vluchttijd is voor 9 uur 's ochtends, dan is de kans op vluchtvertraging 10%*. We passen de door ons ontwikkelde methode toe op een casestudy om het risico van patiënten op de intensive care van een ziekenhuis te voorspellen om opnieuw te worden opgenomen nadat ze zijn ontslagen, waarmee we onder andere laten zien dat het model kan leren van feedback van medische experts—een pilotstudie naar mens-AI samenwerking.

Verder onderzoeken we de taak van discretisatie en samenvatten van tweedimensionale datasets, met als potentieel gebruiksscenario het samenvatten van bestemmingen van trajecten geregistreerd door GPS-apparaten. De onderzoeksvraag hierbij is hoe datasets van geregistreerde locaties te partitioneren met een adaptieve granulariteit. Dit moet noch te grof noch te verfijnd, aangezien het eerste betekent dat een grote hoeveelheid informatie verloren gaat, terwijl het laatste betekent dat ruis in plaats van patronen in de gegevens worden ontdekt.

Tot slot bestuderen we het probleem van het begrijpen van de structuur van (conditionele) afhankelijkheden in gegevens. Dit gaat over het ontwikkelen van algoritmen die op basis van data automatisch conclusies kunnen trekken zoals 'het risico op bosbrand is onafhankelijk van de maand, geconditioneerd op de temperatuur en vochtigheid uit gegevens' (d.w.z., zodra de temperatuur en vochtigheid bekend zijn, is het risico op bosbrand hetzelfde ongeacht de maand van het jaar). We doen zowel theoretisch als methodologisch onderzoek voor een uitdagend gegevenstype, namelijk de mix van discrete en continue waarden. Historische gegevens van bosbranden worden bijvoorbeeld vaak op deze manier geregistreerd, waar de data zowel 'geen brand' (een discreet label) als een oppervlakte van een brand (een kwantitatieve, continue waarde) bevat.

Samengevat, op partitie-gebaseerde modellen voor verschillende taken worden bestudeerd, met als doel ze interpreteerbaarder en transparanter te maken voor de eindgebruiker.