

## Abstract

Occupant behavior (OB) is described as occupants' presence and actions, which influence energy consumption in buildings. OB is not easy to quantitatively describe due to its stochastic, dynamic, multifaceted, and inter-disciplinary nature. OB can be addressed by occupant-centered design and control applications carried out at divergent phases of the building life-cycle, which are sometimes oversimplified. Thus, it is essential to understand OB extensively by consolidating qualitative and quantitative approaches based on machine learning (ML) techniques and pertinent data to design such applications.

The emergence of ubiquitous computing, reasonably priced sensing modalities, and accessibility to high computational power resources have shaped the very essence of occupant behavior estimation and modeling in buildings. One driving force is the use of data to steer critical decisions for design and control applications in buildings. However, the procurement of labeled data is often an expensive process. It is challenging to acquire large-enough datasets to train a supervised ML model efficiently. The imbalanced features of OB data and the applicability of ML models to various buildings are defined as the main constraints of the existing OB modeling practices, which are further elaborated and discussed in the scope of this dissertation.

This overall contribution of this dissertation is to quantify and measure OB using heterogeneous sensing modalities that can be installed in actual buildings in variegated settings. The aim is to create models for more accurate modeling and estimation of OB. OB's seven different resolutions are presence, count, location, social groupings, trajectory, identity, and actions. The main elements for administering OB studies and predicting a targeted resolution depend on the choice of sensing modalities, selecting an appropriate ML model for an intended prediction task, and model complexity. The research is extensively carried out by broadening the scalability, scope and comprehensiveness of sensor data streams, sources, data recording strategies, mining, and pre-processing data. Moreover, annotate and augment the datasets accumulated in heterogeneous case settings.

The data-driven predictive models on the different resolutions of OB obtained using the methodologies proposed in this dissertation allow for more informed decision-making compared to conventional OB modeling methods. In addition, the potential risks are identified in the accumulated datasets, and attempts are made to anonymize the sensitive attributes by assessing the risks and implementing data suppression techniques. Finally, recommending strategies and directions for future research efforts on developing an ontology, enabling data share and reuse for building-related datasets.

## Resumé

Menneskelig adfærd (MA) beskrives som menneskelig tilstedeværelse og dets handlinger, hvilket påvirker energiforbruget i en bygning. MA er svært at beskrive kvantitativt på grund af de stokastiske, alsidige, multidimensionale og tværfaglige karakteristika. MA kan understøttes ved at bruge menneskecentreret design og kontrolapplikationer på forskellige stadier i bygningens livscyklus. Dette bliver til tider forsimplet. Derfor er det nødvendigt at forstå MA dybdegående og konsolidere kvalitative og kvantitative metoder baseret på data og maskinlæring (ML) til at designe sådanne løsninger.

Udbredelsen af allestedsnærværende IT, billige sensor modaliteter, og tilgængeligheden af computer regnekraft har skabt mulighederne for at lave MA estimater og modeller i bygninger. En af drivkræfterne er brugen af data til styring af vigtige beslutninger for design- og kontrolapplikationer i bygninger. ML-algoritmer, som lærer af og forudsiger MA-data, kan implementeres. Imidlertid er det ofte en dyr proces at anskaffe mærket data. Det er udfordrende at tilegne sig store nok datasæt til effektivt at træne en overvåget ML-model. For at afhjælpe dette problem, vil vi i denne afhandling forsøge at mindske problemstillingerne med ubalanceret data.

Denne afhandling bidrager til at kvantificere og måle MA ved hjælp af heterogene sensor modaliteter, der kan installeres i faktiske bygninger i varierede miljøer. Målet er at skabe modeller til mere præcis modellering og estimering af OB. MA's syv forskellige aspekter er tilstedeværelse, antal, placering, sociale grupperinger, kurs, identitet, og handlinger. Hovedelementerne ved administration af MA-studier og forudsigelse af en målrettet løsning afhænger af valget af sensor modaliteter, valget af en passende ML-model til den givne forudsigelsesopgave, og modelkompleksiteten. Forskningen udføres i vid udstrækning ved at udvide omfanget af forskellige datastrømme, kilder, informationsregistreringssystemer, forbehandling af data, afbøde ubalance i data, sensorfusionsstrategier og udvælgelse af passende ML-teknikker. Yderligere, vil vi manuelt mærke og forøge kvaliteten af datasæt akkumuleret i heterogene forsøgsscenarier.

De datadrevne forudsigelsesmodeller på de forskellige løsninger til MA, opnået ved hjælp af metoderne foreslået i denne afhandling, tillader mere informeret beslutningstagning i sammenligning med konventionelle metoder til MA-modellering. Derudover identificerede vi de potentielle risici i de akkumulerede datasæt og forsøgte at anonymisere de følsomme attributter ved at vurdere risici og implementere dataslørningstekniker. Endelig, anbefaling af strategier og retninger til fremtidig forskningsindsats for at udvikle en ontologi, der muliggør datadeling og genbrug til bygningsrelaterede datasæt.