

# マルチモーダルセンサとLLMを用いた 個人の体感推定手法の検討

松田 裕貴<sup>1,2,3,a)</sup> 立花 巧樹<sup>4</sup> 呉 健朗<sup>4</sup> 諏訪 博彦<sup>1,3,b)</sup>

**概要:** 都市規模のデジタルツインの実現に向けては、物理的に計測可能な客観的なデータだけでなく、人々が「どのように感じるか」という主観的な情報も重要となると考えられる。しかしながら、環境や個人による差異が大きいことから、主観的な情報の推定は用意ではない。そこで本稿では、マルチモーダルなセンサデータとLLMを活用し、環境や個人差を吸収した個人の体感推定手法について検討する。

## Toward Individual Perception Estimation using Multimodal Sensor and LLM

### 1. はじめに

Society 5.0 の時代を見据え、我々の住む現実世界 (Physical Space) をデジタル世界 (Cyber Space) に写し取ることにより、様々な分析・シミュレーションを可能とする都市規模のデジタルツイン (Digital Twin) の実現が期待されている。近年のセンシング技術の発展に伴い、物理的に計測可能なものを中心として都市空間における様々な事象がセンシング可能となりつつあるが、未だセンシングできていない事象も数多く残されている。

著者らは、物理的に計測可能な事象だけではなく「人がどのように「事象」を観測しているのか?」という観点に着目している。例えば、ある空間での「混雑度」とは、一般的には「物理的な人の数」を指すことが多いだろう。しかしながら、実際に「混雑度」の情報を利用してサービスを提供する場合を考えると、物理的な人数 (=客観的な混雑度) だけではなく、人が「どの程度混雑していると「感じる」か?」という情報 (=主観的な混雑度) も重要である可能性がある。同様に、気温であれば「物理的な温度」に加えて「体感温度」、時間であれば「実際に掛かった時間」に加えて「体感時間」といったように、「人が観測した世界」という観点から事象をセンシングすることで、より人の主観・体感に合わせたサービスが提供できる可能性がある。

「体感〇〇」を推定する手法については、これまでの研究で数多く取り組まれている。例えば、体感温度について

は空調製品に組み込まれるなど活用事例が存在する。一方で、体感混雑度など、多様な環境・多様な利用者への対応が求められる尺度については、モデルを一般化することが難しいことが課題となる。そこで本稿では、マルチモーダルなセンサで得られたデータを環境・個人の差異を吸収した形で取り扱うために、LLM (Large Language Model, 大規模言語モデル) を導入し、個人の体感推定手法を実現する方法について検討する。

### 2. 関連研究

「体感温度」や「体感音量」は、日常生活でも差を体感しやすいことから、それらを対象にした研究もいくつか存在する。三輪らは、室内に居る人の体感温度を、熱画像データから推定される部屋形状や輻射熱、人の姿勢や人自身の温度といった情報を用いて推定し空調設備を自動制御する手法を提案した [1]。三栖らは、室内の温度が一定であったとしても照明の色によって体感温度が変化することを明らかにした [2]。寺田は、ある音を提示した時の体感音量が、提示前・提示中に聞こえる周辺の雑音によって変化することを明らかにし、この影響を考慮した上で正確な情報を人に伝える音情報提示システムを提案した [3]。小林らは、VR 空間内における映像が聴覚に与える影響について調査し、視覚刺激によって体感音量が変化することを明らかにした [4]。

人間の五感だけでなく、人が事象をどのように解釈するのかについても体感による影響がある。著者らは、クラウドソーシングによる調査によって「体感混雑度」が実際の人数 (客観的な混雑度) と必ずしも一致せず、人によるばらつきが生じることを明らかにした [5]。中村らは、照明刺激を変化させた際の「くつろぎ感」が、心理的指標および生理的指標 (脳波や心電) とどのように対応付けられる

<sup>1</sup> 奈良先端科学技術大学院大学,  
Nara Institute of Science and Technology

<sup>2</sup> 岡山大学, Okayama University

<sup>3</sup> 理化学研究所, RIKEN AIP

<sup>4</sup> ソフトバンク株式会社, SoftBank Corp.

a) yukimat@is.naist.jp

b) h-suwa@is.naist.jp

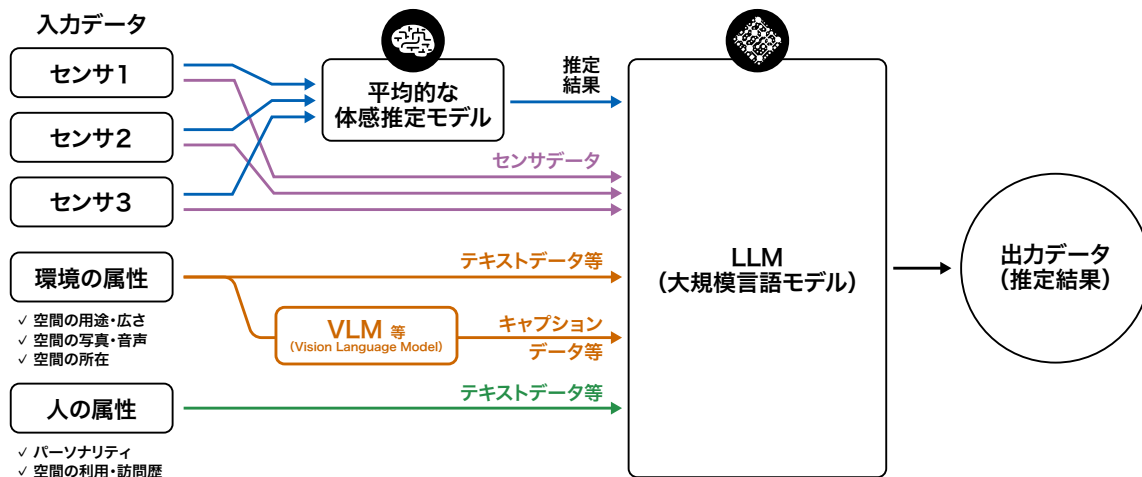


図 1: マルチモーダルセンサと LLM を用いた個人の体感推定のアプローチ

のかを明らかにした [6]. 飛谷らは, CG で構築した人の素肌の質感が視覚的にどのような印象を与えるのかの調査を通じて, 素肌の物理特性情報から人に与える印象を推定可能にした [7]. 河中らは, ドライブ中の景色の「観光地らしさ」がどのような印象によって決定づけられるのかを調査し, 観光地らしさを構成する尺度を明らかにし, さらにドライブレコーダによって撮影された映像から, それらの尺度に基づくスコアを推定するモデルを構築した [8].

### 3. 提案手法

ここでは, マルチモーダルなセンサで得られたデータを環境・個人の差異を吸収した形で取り扱い, 個人の体感推定手法を実現するために, LLM を活用する方法を提案する. 以降では, 「体感混雑度」を具体的な推定タスクと設定し, 議論を行う.

アプローチを図式化したものを図 1 に示す. まず, マルチモーダルセンサデータに基づいて平均的な環境・個人の体感を推定するベースモデルを構築する. このベースモデルの出力結果は当然ながら環境・個人に特化したものではないため, 入力データ (センサデータ), 出力結果, 環境・個人の属性情報を LLM に与えることで出力をフィッティング, 環境・個人に合わせたものとする. 環境の属性情報については, 例えば空間の用途や広さなどといったパラメータとして与えられるものから, 空間の写真・音声といった高次元データ (Vision Language Model によってゼロショットで写真をキャプション・言語化し, 入力として与える等), 位置情報といった周囲の環境との相対関係を示すデータ (地図検索などと組み合わせ, 周囲の土地・空間利用状況などを入力として与える等) など, 様々な方法が考えられる. また, 個人の属性情報については, Big Five Personality Traits といった個人のパーソナリティを示すデータや, これまでの空間の利用歴・訪問歴などといった情報を与えることが考えられる. 最終的に, センサデータおよび環境・人の属性に関する情報を LLM に事前

情報として与え, 平均的な体感推定モデルの出力がどの程度対象ユーザにおいて変動するのかを LLM に推論させることにより, パーソナライズした体感推定結果を得る.

### 4. おわりに

本稿では, マルチモーダルなセンサデータと LLM を活用し, 環境や個人差を吸収した個人の体感推定手法について検討を行った. 今後は具体的なユースケースを対象に提案手法を実装し, その有用性について検証を行う.

### 参考文献

- [1] 三輪祥太郎, 渡邊信太郎, 平位隆史, 松本崇. 体感温度を制御するエアコン. 日本ロボット学会誌, Vol. 32, No. 3, pp. 218–221, 2014.
- [2] 三栖貴行, 小田原健雄, 渡部智樹, 一色正男. LED 照明の光色変化による心理的影響と体感温度の変化. 日本色彩学会誌, Vol. 42, No. 3+, pp. 205–208, 2018.
- [3] 寺田努. 主観的コンテキストに基づく情報提示システム. マルチメディア, 分散, 協調とモバイルシンポジウム (DICOMO2012) 論文集, pp. 737–743, 2012.
- [4] 小林永実, 松田裕貴, 渡邊拓貴, 安本慶一. VR 空間の単純な視覚刺激が及ぼす体感音量への影響調査. マルチメディア, 分散, 協調とモバイル (DICOMO 2023) シンポジウム論文集, pp. 1501–1505, 2023.
- [5] Yuki Matsuda. IoPT: A Concept of Internet of Perception-aware Things. In *The 12th International Conference on the Internet of Things (IoT '22)*, pp. 201–204, 2022.
- [6] 中村透, 上垣百合子, 藤原ゆり, 奥谷晃久, 山本松樹, 長田典子. 照明刺激環境下のくつろぎ感に関する心理生理学的研究. 電子情報通信学会論文誌 D, Vol. 96, No. 6, pp. 1536–1544, 2013.
- [7] 飛谷謙介, 松本達也, 谿雄祐, 藤井宏樹, 長田典子. 素肌の質感表現における印象と物理特性の関係性のモデル化. 映像情報メディア学会誌, Vol. 71, No. 11, pp. J259–J268, 2017.
- [8] Masaki Kawanaka, Yuki Matsuda, Hirohiko Suwa, and Keiichi Yasumoto. Dashcam Video Curation for Generating Memorial Movies on Tourism using Multiple Measures of “Tourist Spot Likeness”. In *The 25th International Conference on Human-Computer Interaction (HCHI '23)*, pp. 238–249, 2023.