Leveraging Human Pose Estimation to improve Health Recommender Systems

Gaetano Dibenedetto

University of Bari Aldo Moro - Department of Computer Science, Via Orabona 4, Bari, 70125, Italy

Abstract

In recent years, the integration of multimodal and multi-source data has gained attention for its potential to enhance the accuracy and relevance of recommender systems. While Health Recommender Systems (HRS) predominantly rely on patient-specific data, the inclusion of Pose Estimation (PE) data remains unexplored.

My Ph.D. research aims to bridge this gap by investigating and incorporating PE as a data source within HRS. This will also focus on addressing critical challenges such as ensuring user privacy and optimizing the trade-off between system performance and real-time responsiveness.

Keywords

Health Recommender Systems, Human Pose Estimation, Privacy, Explainability

1. Introduction

This widespread use of wearable health devices and online health information systems has generated a growing user need for more personalized health advice, a challenge that Health Recommender Systems (HRS) aims to address. Despite their potential, current HRSs face the challenge of aligning their recommendations with users' expectations, a key factor in building trust in such systems. HRS found a strong synergy with Human Pose Estimation (HPE). Indeed, observing poses risky to the user's health can be crucial in providing effective recommendations and support and studies applied to the healthcare domain. For example, HPE is adopted in the field of occupational medicine for conducting ergonomic postural assessment. Indeed, one of the primary reasons for nonattendance from work is the health problems stemming from repeated improper postures and movements [1]. To address these issues, ergonomists assess the posture through direct on-site observation or by analyzing video recordings of workers performing their routine job tasks. Traditional postural assessment methods often rely on standardized indices that provide scores based on evaluations of various aspects, such as physiological body angles, load weight, and number of repetitions. In HRS, a possible innovative aspect is to leverage data gathered from HPE techniques not only to enhance performance but also to provide more personalized explanations based on both the user's characteristics and actions. Building on these ideas, I proposed a preliminary work [2] focused on posture correction for office workers. In the literature, many studies share our goal of posture classification [3, 4, 5], but their approaches rely on data collected under strict constraints such as the use specialized cameras, sensors or other devices embedded into chairs. In contrast, I have proposed a simple approach based on data collected from classical cameras and lightweight, fast AI-based classification models. By analyzing the results of the classification model, we are then able to suggest corrections to the pose for improving worker well-being. The goal I am therefore committed to pursuing during my Ph.D. is to introduce a novel approach that integrates data from HPE into HRS, paving the way for more precise and personalized recommendations.

Doctoral Consortium at the 23rd International Conference of the Italian Association for Artificial Intelligence, Bolzano, Italy, November 25-28, 2024

gaetano.dibenedetto@uniba.it (G. Dibenedetto)

https://linkedin.com/in/gaetano-dibenedetto/ (G. Dibenedetto)

0000-0001-6083-3600 (G. Dibenedetto)

2. State Of The Art

2.1. Health Recommender Systems

HRS suggest personalized health information to individuals based on their specific needs and preferences. These systems goal is to improve health outcomes by delivering relevant, accurate, and up-to-date health information, while reducing the time and cost associated with the decision-making process.
Given the interplay between Recommender System (RS) and the exploration of HPE, our focus will be on *Health Status Prediction Systems* and *Physical Activity Recommender System*. Health Status Prediction Systems (HSPS) adopt machine learning algorithms to understand complex relationships between self-reported health concerns and their outcomes, allowing them to predict users' current health status. These systems are particularly beneficial for older patients and individuals with pre-existing conditions, as they integrate data from wearable sensors to provide real-time monitoring and alert users to potential health issues. As an example, in [6], a smart HSPS for predicting hypertension and type 2 diabetes is presented. This system collects health-related data, such as blood pressure, weight, and physical activity, via wearable devices and home-based sensors. The data is then processed and analyzed using serveral machine learning algorithms (e.g., decision trees, random forests, neural networks) to predict the probablity of developing these conditions.
Physical Activity Recommender Systems (PARS) focus on the user's current health status and demographic factors like age and gender to suggest personalized daily exercise routines. These systems, often integrated into wearable devices, continuously collect user data such as calories burned, daily step count, and heart rate. In [7], a PARS designed for patients with arterial hypertension, a condition characterized by high blood pressure, is presented. The system uses data from wearable devices, such as heart rate monitors, to track physical activity levels and provide recommendations based on the patient's age, gender, and health status.
Although the extensive literature on HSPS and PARS is extensive and relevant to understanding how the scientific community is moving on the topic, **as far I know, no work incorporates Human Pose Estimation into HRS, i.e., the primary focus of my Ph.D. research.**

2.2. Human Pose Estimation

A key aspect of my Ph.D. studies is about how efficiently estimate Human Pose. *Human Pose Estimation* (HPE) is a well-established field in computer vision that focuses on predicting human body parts based on the analysis of images and videos. The rapid advancements in deep learning have demonstrated its superiority over traditional computer vision techniques in tasks such as image classification [8], semantic segmentation [9], and object detection [10]. In literature there are both 2D and 3D approaches for HPE. 2D systems focus to identify and tracking body keypoints in two-dimensional images or videos. These methods are computationally efficient and capable of providing real-time results; however, they may be less accurate than 3D methods, especially in complex poses or cases involving occlusion. For 2D HPE, the best results today are achieved by transformer-based models. Transformer architectures have recently gained prominence and have been shown to be effective in HPE. According to benchmark results on the COCO Dataset [11], the leading model is ViTPose [12]. It employs a plain and non-hierarchical vision transformers as backbones to extract features for a given person instance and a lightweight decoder for HPE. It is highly scalable, with model sizes ranging from 100M to 1B parameters, taking advantage of the transformer's capacity for scalability and parallelism. In the context of 2D-to-3D lifting approaches, 3D HPE inspired by recent advances in 2D HPE has gained popularity as a solution. By leveraging the strong performance of 2D pose detectors, 2D-to-3D lifting approaches generally outperform direct 3D estimation methods [13]. A notable transformer-based architecture in this domain is MotionBERT, which achieves state-of-the-art results on the Human3.6M dataset [14]. MotionBERT incorporates a motion encoder during pretraining to reconstruct 3D motion from incomplete 2D observations by integrating geometric, kinematic, and physical insights into human movement. **By considering such research directions, I will continue to investigate approaches Transformer based.**

3. RESEARCH APPROACH, METHODS, AND RATIONALE

The main objective of my PhD project is to design and **develop an HRS that includes HPE data**. In the following, I present possible pipeline starting from gather this data and possible strategies for a possible project of mine that integrate HPE into a HRS, and how to evaluate them.

Recording Data. The scarcity of visual data related to the medical sector, linked to health data such as medical records, is due to the difficulty privacy restrictions. One potential solution to address these restrictions is to implement real-time adjustments at the recording source, making it easier to obtain new data within the healthcare environment. Such modifications could include:

- **Real-time preprocessing of the videos or images**. As demonstrated in [15], facial blurring effectively safeguards privacy without causing a statistically significant difference on kinematic calculations. In the context of HRS this approach could allow systems to gather more informative data for performance enhancement or detailed explanations based on specific actions. Incorporating activity data from HPE can provide a more comprehensive and understandable experience, also for non-expert users in the domain.
- **Storing only data derived from HPE models instead of original image or video sources**. Recent advancements in HPE models have shown remarkable results, with scalable architectures that offer flexibility in selecting the most suitable option for real-time inference. This enables camera-based systems to achieve high accuracy without the need for powerful servers. Studies, such as [16], have demonstrated that deep neural networks processing only keypoints can perform comparably, or sometimes even better, than models relying on original sources alone. Moreover, training a model based on a limited number of keypoints is more efficient than training one from the large volume of pixels in each frame.

Strategies to include HPE data into HRS. A possible approach to be used for integrating HPE into HRS is to treat the pose as a categorical descriptive feature. Therefore, each pose is classified into predefined categories, enabling the RS to analyze patterns and generate recommendation based on these categories, such as human action recognition task [17]. For instance, poses indicating different levels of physical activity can be classified as sedentary, moderate, or vigorous, helping to provide personalized health recommendations. Another approach involves representing pose data as a vector of keypoints or as a graph embedding. Each detected keypoint or graph node corresponds to a specific body part or joint, and their spatial relationships are captured to create a comprehensive pose representation. This representation can then be integrated with other data sources in a multi-source setting, as seen in [18]. By using this method, the RS can better understand body posture and movement dynamics, enabling more personalized and context-aware recommendations.

Performance Evaluation and Considerations. A key aspect of this project is the evaluation of the trade-off between the pose detection accuracy and the responsiveness of the recommender system. High-quality models can yield more precise recommendations but may compromise computational efficiency, making them unsuitable for real-time environments. Therefore, it is crucial to find the right balance between accuracy and responsiveness to meet the specific requirements of the application. This balance will be explored in the context of a real-world scenario. Since my scholarship is supported by a healthcare-affiliated company, there is a promising opportunity to develop and test a prototype of this system in a practical setting. Additionally, the system will include an explanatory module that allows users to understand the reasoning behind specific recommendations, enhancing both user engagement and comprehension. In the context of active aging, pose detection and wearable devices can play a vital role in combating sedentary lifestyles. By continuously monitoring an individual's routine and posture, the system can identify inactivity patterns and offer personalized interventions to encourage physical activity and improve overall health.

4. RESEARCH QUESTIONS AND ONGOING WORKS

To assess the effectiveness of the proposed methodology, the following research questions are posed:

RQ1 - *Does a specific HPE method yield better results than others?*

Several state-of-the-art HPE techniques exist (Section 2.2), ranging from 2D to 3D PE, with sub-categories such as Single-Person, Multi-Person, Top-Down, and Bottom-Up approaches. Choosing the best HPE method to integrate into HRS depends on a various factors, such as the environment, data availability, and contextual requirements. To better understand the strengths and limitations of different HPE methods, I conducted a comprehensive literature review on pose estimation techniques, comparing methods for 2D/3D pose detection from images and videos across different scenarios (e.g., single/multi-person, monocular/multi-view input). This review, summarized in the paper "Comparing Human Pose Estimation through Deep Learning Approaches: An Overview", currently under submission to Elsevier Computer Vision and Image Understanding, will guide the selection of the most appropriate technique for the real-time environment I am working on. I gained practical experience with HPE methods through the development of a posture correction system for office workers [2].

RQ2 - *Are there available HPE data for the developing of my work?*

Section 2.1 highlights a lack of existing work incorporating HPE data. As my scholarship is funded by a specialized company working on healthcare applications and HPE, i.e. Naps Lab S.r.l.s., eases the process of defining an environment for working on real-world HRS, where pose estimation data can be collected, thus addressing the issue of data scarcity. Nowadays, healthcare facilities, even those for elderly people, are technologically well-equipped, making easier obtaining this type of data from a hardware perspective. In order to start my work on it, while working on [2], I have already created a new dataset capturing data from a webcam, which is currently available on Zenodo.

RQ3 - *How can HPE data be incorporated into a recommender system?*

At present, we do not yet have a clear idea of how HPE data can be integrated into a successful HRS. We are investigating the idea of using context-based recommender systems models based on Transformer architectures, where the visual data regarding the subject's location is provided directly as input to the model. In this way, the output of the recommender system can be an item from any domain where a correlation with human posture can be observed, e.g., physical exercises. This point, however, is the one on which a discussion with the AIxIA community may be most helpful in successfully addressing my future studies.

RQ4 - *How can I measure the performances of the model?*

With limited directly comparable prior research, it is critical to define appropriate evaluation methodologies. I will evaluate single modules in offline mode by using specialized benchmarks and datasets. An optimal solution to evaluate the whole architecture will be to obtain feedback from experts in the healthcare field. Thanks to Naps Lab S.r.l.s.'s expertise, I will have the opportunity to evaluate the system not only in-vitro but also in real-world (in-vivo) environments. This will involve comparing HRS performance using only medical data against systems that integrate HPE data as well. Feedback from users and medical professionals will be instrumental in refining and improving the research.

RQ5 - *How how to deal with privacy and trustworthy?*

Privacy management is extremely important, particularly in the healthcare sector. The collaboration with Naps Lab S.r.l.s. will enable the collection of real data, but at the same time makes it mandatory to implement techniques to protect personal and medical data. This is a very critical aspect of my work, which cannot be overlooked. Other researchers experiences and suggestions on the topic will be strongly appreciated. As stated in [19], *the necessity to be able to explain the technology used for medical decision-making processes represents a normative standard unquestioned in its principle relevance.* In my work on posture correction system for offices [2], the dataset have been published on Zenodo[1], without including any original video or image data that might raise privacy concerns, but this can strongly affect the replicability of proposed approaches.

[1] https://zenodo.org/records/11075018

5. LONG-TERM GOALS

As long term goals, I aim to advance the actual state of the art HPE models by exploring different novel architectures based on Foundational Large Multimodal Models like META-Sapiens [20]. While I already explored the use of ViTPose [12], a state-of-the-art 2D HPE known for its high accuracy and scalability, allowing for configurations ranging from 100 million to 1 billion parameters, a further solution could be the use of 3D HPE, e.g. MotionBert [21], providing more comprehensive measure for posture comparison. In the current work, a simulation of the Epipolar Geometry computation have been carried out, to compare different postures taken from different viewpoints. While effective, it may not be as precise as a true 3D rotation-based approach.

Additionally, I plan to expand the Health Recommender System [2], which is currently able to rank poses and suggest corrections to users. My goal is to extend this functionality to detect potential long-term postural deviations such as scoliosis, kyphosis, and lordosis. Based on these detection, the system could recommend a nearby specialist for targeted correction.

A possible recommendation could result from the detecting categorical anomalies, such as imbalances in the shoulders, hips, or head. If these posture misalignments are identified as long-term issues, they may affect physical appearance. Preventative measures, like targeted physical exercises designed to strengthen specific muscles, could help address these imbalances. Recommendations for such exercises could be presented as a list of items within a recommender system, with detailed descriptions generated by a large language model, making the exercises accessible to non-expert users and promoting regular engagement. Additionally, if the user already performs some of these exercises during personal training, the system could incorporate contextual information, such as daily workout routines, to ensure personalized recommendations that avoid redundancy.

6. Conclusion

I started my Ph.D. in October 2023 at University of Bari Aldo Moro, under the supervision of Prof. Pasquale Lops and Dr. Marco Polignano, belonging to the SWAP research group[2], as well as Dr. Giuseppe Cavallo from Naps Lab S.r.l.s.[3]. I expect to complete my Ph.D. at the end of 2026. After almost a year of working on the topic, I am at the stage where I am beginning to approach operational techniques and strategies for achieving successful HRS. This stage is the crucial one for the success of my journey, and therefore, I believe that I can benefit greatly from discussion and valuable suggestions from experts in Computer Vision, Recommender Systems, and eHealth and, in particular, from the vibrant AIxIA community. I would, therefore, be delighted to attend the conference and the Doctoral Consortium.

Acknowledgments

This research is partially funded by PNRR - Mission 4 ("Education and research") – Component 2 ("From research to business"), Investment 3.3 ("Introduction of innovative doctorates that respond to the innovation needs of companies and promote the hiring of researchers by companies") D.M.n. 117/2023 - H91I23000170007.
I extend my sincere gratitude to Naps Lab S.r.l.s.[3] for their support and collaboration in the realisation of this research.

References

[1] L. Punnett, D. H. Wegman, Work-related musculoskeletal disorders: the epidemiologic evidence and the debate, Journal of electromyography and kinesiology 14 (2004) 13–23.

[2] https://swap.di.uniba.it/
[3] www.napslab.it

[2] G. Dibenedetto, M. Polignano, P. Lops, G. Semeraro, Human pose estimation for explainable corrective feedbacks in office spaces, in: Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization, 2024, pp. 264–275.
[3] S.-H. Han, H.-G. Kim, H.-J. Choi, Rehabilitation posture correction using deep neural network, in: 2017 IEEE international conference on big data and smart computing (BigComp), IEEE, 2017, pp. 400–402.
[4] J. C. T. Mallare, D. F. G. Pineda, G. M. Trinidad, R. D. Serafica, J. B. K. Villanueva, A. R. D. Cruz, R. R. P. Vicerra, K. K. D. Serrano, E. A. Roxas, Sitting posture assessment using computer vision, in: 2017IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM), IEEE, 2017, pp. 1–5.
[5] Y. M. Kim, Y. Son, W. Kim, B. Jin, M. H. Yun, Classification of children's sitting postures using machine learning algorithms, Applied Sciences 8 (2018) 1280.
[6] S. P. Chatrati, G. Hossain, A. Goyal, A. Bhan, S. Bhattacharya, D. Gaurav, S. M. Tiwari, Smart home health monitoring system for predicting type 2 diabetes and hypertension, Journal of King Saud University-Computer and Information Sciences 34 (2022) 862–870.
[7] L. R. Ferretto, E. A. Bellei, D. Biduski, L. C. P. Bin, M. M. Moro, C. R. Cervi, A. C. B. De Marchi, A physical activity recommender system for patients with arterial hypertension, IEEE Access 8 (2020) 61656–61664.
[8] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems 25 (2012).
[9] J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 3431–3440.
[10] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, Advances in neural information processing systems 28 (2015).
[11] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft coco: Common objects in context, in: Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13, Springer, 2014, pp. 740–755.
[12] Y. Xu, J. Zhang, Q. Zhang, D. Tao, Vitpose: Simple vision transformer baselines for human pose estimation, Advances in Neural Information Processing Systems 35 (2022) 38571–38584.
[13] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, M. Shah, Deep learning-based human pose estimation: A survey, ACM Computing Surveys 56 (2023) 1–37.
[14] C. Ionescu, D. Papava, V. Olaru, C. Sminchisescu, Human3. 6m: Large scale datasets and predictive methods for 3d human sensing in natural environments, IEEE transactions on pattern analysis and machine intelligence 36 (2013) 1325–1339.
[15] J. Jiang, W. Skalli, A. Siadat, L. Gajny, Effect of face blurring on human pose estimation: Ensuring subject privacy for medical and occupational health applications, Sensors 22 (2022) 9376.
[16] R. Hachiuma, F. Sato, T. Sekii, Unified keypoint-based action recognition framework via structured keypoint pooling, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2023, pp. 22962–22971.
[17] H. Mollaei, M. M. Sepehri, T. Khatibi, Patient's actions recognition in hospital's recovery department based on rgb-d dataset, Multimedia Tools and Applications 82 (2023) 24127–24154.
[18] G. Spillo, C. Musto, M. Polignano, P. Lops, M. de Gemmis, G. Semeraro, Combining graph neural networks and sentence encoders for knowledge-aware recommendations, in: Proceedings of the 31st ACM UMAP Conference, 2023, pp. 1–12.
[19] H. Kempt, N. Freyer, S. K. Nagel, Justice and the normative standards of explainability in healthcare, Philosophy & Technology 35 (2022) 100.
[20] R. Khirodkar, T. Bagautdinov, J. Martinez, S. Zhaoen, A. James, P. Selednik, S. Anderson, S. Saito, Sapiens: Foundation for human vision models, 2024. URL: https://arxiv.org/abs/2408.12569. arXiv:2408.12569.
[21] W. Zhu, X. Ma, Z. Liu, L. Liu, W. Wu, Y. Wang, Motionbert: A unified perspective on learning human motion representations, in: Proceedings of the IEEE/CVF, 2023, pp. 15085–15099.