Structural Learning in the design of Perspective-Aware AI Systems using Knowledge Graphs

Marjan Alirezaie[1], **Hossein Rahnama**[2,3], **Alex Pentland**[2]

[1]Flybits Labs., TMU Creative AI Hub, Toronto, Canada
[2]MIT Media Lab, Massachusetts Institute of Technology, USA
[3]Toronto Metropolitan University, Toronto, Canada

Abstract

We define perspective-aware computing as an emerging area of computational innovation in which users of the system can view and interact through each other's points of view without the need for a centralized recommendation system. To achieve this, we propose a multi-modal neurosymbolic graph generation approach to construct personalized models known as "borrowable identities" from a user's digital footprint, comprehending an individual's cognitive and behavioral tendencies in diverse and contexts. Applications of our approach enable users of a trusted social network to view and interact with information through each other's perspective. In summary, we allow individuals to lend their expertise to each other, and advance classic digital personalization techniques toward more participatory systems. This approach has potential in the design of less-biased recommendation systems in areas such as Digital Immortality, peer-to-peer learning, and in general, decentralized computational social systems.

Introduction

The concept of user modeling in the realm of human-computer interaction (HCI) has been synonymous with the process of collecting data from users, their preferences, and providing more personalized and context-aware experiences (Fischer 2001). In this classic approach, the source user, from whom the data is collected, is either identical or shares similarities with the target user that the machine aims to serve. However, by shifting the user modeling approach, we can now distinguish between the source and the target users, thereby expanding the spectrum of HCI applications towards scenarios relying on what we call "borrowable identities" and "Through-Perspective Computing". By borrowable identities and In-Perspective Computing, we refer to a digital representation reflecting one's identity and behavioral patterns that can be shared with others. This approach enables individuals to perceive or even interact through the unique lenses of each other with different viewpoints. These lenses are essentially well-encapsulated knowledge-based models in which their ontologies can dynamically adapt based on the context of usage. It could also advance research areas such as peer-to-peer learning, biased-reducing systems, and emerging concepts such as digital immortality that aim to extend a person's digital presence and facilitate conversations from their viewpoint in their physical absence. This paradigm shift in creating user models offers humanity the opportunity to gain valuable insights into how reality is perceived from various perspectives, fostering understanding and enriching not only human-machine, but also human-human interactions.

Many user modeling approaches tend to focus on specific data capture, such as analyzing user motion within well-defined contexts like cyber-physical settings or interactions with particular systems. These approaches often target a specific user category within a defined environment, lacking adaptability for broader use cases (Anders et al. 2022). To create a borrowable or digital identity as a computational model of the user, an algorithm must go beyond simply understanding the user's preferences in one specific domain, and instead acquire knowledge of their cognitive processes and behavioral patterns across domains, effectively capturing the knowledge of their mentality and personality (Nersessian 1992; Treur and van Ments 2022). In today's digital age, as social media platforms become an integral part of our lives, our online presence leaves an increasingly significant digital footprint. Every photo, status update, tweet, or song we engage with paints a picture of our interests, relationships, and evolving personas that can continuously contribute to a model, gradually revealing facets of our personality. The objective is to systematically leverage the extensive repository of any available digital footprint, augmenting it by incorporating established theories from psychology, social science, and the humanities. This endeavor aims to craft a digital representation, as a result of a formal computation, that faithfully captures an individual's cognitive processes, evident in their digital journey, thereby reflecting their cognitive paradigms and behavioral inclinations across various scenarios.

Achieving this ambitious goal necessitates addressing a series of critical challenges.

- **C1:** Efficiently collecting diverse data while respecting user privacy.
- **C2:** Integrating multimodal data for consistent and relevant representations.
- **C3:** Creating a coherent user mental model for informed decision-making.

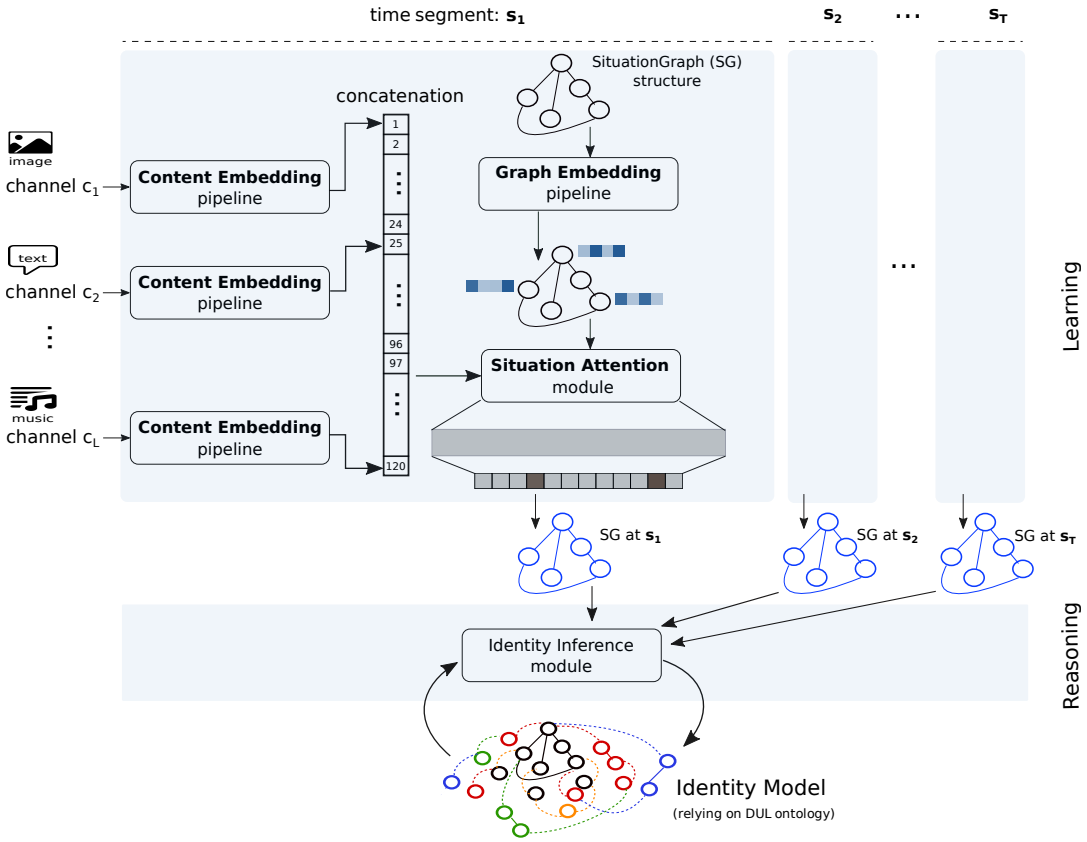

Figure 1: Model architecture overview for creating a digital identity model of an individual using a neuro-symbolic, multi-modal, and knowledge-aware solution

The challenges mentioned above underscore the complexities associated with utilizing digital footprints to build digital mental models of individuals. In this position paper, we propose a workflow founded on neurosymbolic (NeSy) AI solutions, as depicted in Figure 1. This approach relies on the integration of machine learning (for data perception and feature learning) and machine reasoning (for identity inference through the construction of models adhering to rules and constraints) (d'Avila Garcez and Lamb 2020).

In the following sections, we explore our proposed model, specifically a fusion of a multi-modal knowledge-aware graph learning model and a reasoning mechanism designed to tackle challenges **C2** and **C3**. This reasoning mechanism aims to deduce potential causes from observed effects by adhering to established theories and rules (Bochman 2003). We also investigate potential solutions, such as decentralized data models, to mitigate challenge **C1**.

Methodology

Figure 1 provides an overview of the architecture for our proposed model. We assume that the data is sourced from L data channels $\{c_1, c_2, ..., c_L\}$, each associated with a specific data type (e.g., an image, a piece of text, etc.). In the realm of understanding a user's identity and behavioral patterns from digital footprints, considering time and temporal relations among data points is crucial. While pinpointing exact event timestamps may not be necessary, recognizing temporal proximity is crucial. For example, a 15-minute gap between a user's Instagram post and their activity of liking a friend's post on Facebook can still be considered synchronized. To achieve this, we divide a day into T time segments, denoted as $\{s_1, s_2, ..., s_T\}$. If data from different channels fall within the same time segment, we consider them synchronized. The value of T is context-dependent, allowing for flexibility in understanding user behavior.

As shown in Figure 1, our approach utilizes a knowledge-aware neurosymbolic pipeline to address challenges **C2** and **C3**. It achieves this by consolidating content from various sources into a predefined graph structure and transforming it into a ready-to-reason ontology structure. The following provides further technical details of the proposed approach.

Task Formalization:

Data Channels and Embeddings: We assume there are L data channels providing data instances with the specific data types. For instance, channel c_1 may deliver images from the user's Instagram account, while channel c_2 may provide textual information, such as user comments on a tweet. For every piece of data obtained from channel c_i, we utilize an embedding module denoted as e^i, designated for content embedding. This module is tailored to the specific data type,

for instance, pipelines based on pre-trained Vision Transformer (ViT) (Dosovitskiy et al. 2020) for image content or CLIP (Radford et al. 2021) for image-text combination contents. Given the data, the content embedding pipeline generates data embeddings denoted as $e^i_{c_i} \in R^{(N \times D_i)}$, where N is the number of instances and D_i represents the dimensionality of the data embeddings corresponding to channel c_i. As shown in Figure 1, the embeddings received from different modules are concatenated to feed the *Situation Attention* module explained in the following sections.

Graph Embedding: Knowledge awareness within our model implies its ability to be guided by external knowledge sources, thus enhancing the feature extraction and learning processes. To facilitate this, we employ a predefined knowledge graph known as the Situation Graph (SG). The SG serves as a formal representation for defining various situations, encompassing aspects such as ambiance, sentiment, emotions, qualitative time, and location information. It operates as a filtering mechanism, streamlining the extraction of relevant details from diverse data channels while adhering to the predefined structure. We represent the SG knowledge graph as $G(V, E)$, where V consists of nodes such as `Sentiment` (including types like `Positive`, `Negative`, and `Neutral`), `Emotion`, `Location`, and more. These nodes are interconnected by edges denoted in E, signifying properties and relationships. For instance, the triple `(tweet_123, hasSentiment, PositiveSentiment)` is a common example within SG. To learn how to extract a formal representation of situations from content in each channel (an end-to-end process), we employ a module to generate embeddings for the relevant entities within the SG graph. To achieve this, we employ GraphSAGE (Hamilton, Ying, and Leskovec 2017) or node2vec (Grover and Leskovec 2016), which, while not pre-trained, are purpose-built to effectively learn graph structures and transform them into vector representations. Alternatively, we can convert knowledge graph triples, such as those in SG, into sentences and utilize language models like K-BERT (Liu et al. 2020), leveraging BERT (Devlin et al. 2018), to generate graph embeddings. This approach facilitates the development of high-performance models with significantly reduced data requirements.

The embedding resulted from the chosen graph embedding model is represented as $e_{sg} \in R^{(P \times D_{sg})}$, where $P = |V| + |E|$ represents the total number of entities (including nodes and edges) in the graph, and D_{sg} signifies the dimensionality of the graph embeddings.

Situation-Attention Module: The model employs a knowledge attention module, referred to as *Situation Attention*, to align and calculate attention scores between data embeddings and the SG embedding. It takes as input the concatenated data embeddings from all channels, denoted as $e = [e^1_{c_1}, e^2_{c_2}, \ldots, e^L_{c_L}] \in R^{(N \times \Sigma D_i)}$, along with the SG embedding e_{sg}.

The attention scores represented by $A \in R^{(N \times P)}$ (where N is the number of instances and P is the number of entities in the SG), capture the relevance of data embeddings to different entities in the graph. Subsequently, a softmax operation is applied along the rows (instances) to obtain attention weights, denoted as $W \in R^{(N \times P)}$. Each entry $W[i, j]$ in the attention matrix signifies the weight or attention allocated to the SG entity j for the i-th instance in the data. To create a consolidated knowledge-aware representation for each instance, a weighted combination of the SG embedding and attention weights is computed. Specifically, for each instance i, the weighted combination is calculated as follows: $O[i, :] = \Sigma(W[i, j] \times e_{sg}[j, :])$. In this equation, j spans across all SG entities. The formula computes a weighted sum of SG embeddings, where each embedding is weighted by the attention score associated with the corresponding SG entity for the particular instance. This process results in an aggregated representation for each instance, taking into account the relevance of different SG entities based on the computed attention weights.

SG Entity Classification: The attention layer is followed by the entity classification layer to classify the instance features as entities in the SG graph. This layer takes as input the weighted combination of embeddings O. The size of the output layer is equal to the number of entities in the SG graph (P). Applying a softmax activation function to the output of the entity classification layer will convert the aggregated representation into a probability distribution over the SG entities for each instance. The resulting probabilities for each instance feature to be classified as one of the entities in the SG graph represent the output of the softmax layer.

To train the model and ensure the creation of a consistent SG graph populated with content received from the various channels, we employ a loss function that promotes the alignment of predicted graph embeddings with the ground truth representations derived from the channel content, thus guiding the model towards accurate and coherent SG graph construction. Depending on the size and quality of the training dataset, different loss functions may be utilized, ranging from the cross-entropy-based approaches to custom ones specifically designed for the task of graph population. The loss cross-entropy-based loss function measuring the binary classification error for each label or entity independently is defined as follows: where $\hat{Y}_i$ is the predicted probabilities for each entity for the i-th instance, Y_i is the ground truth one-hot encoded labels for each entity for the i-th instance, N is the number of instances, and P is the number of entities in the SG graph:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{P}(Y_i[j].\log(\hat{Y}_i[j])) + ((1 - Y_i[j]).\log(1 - \hat{Y}_i[j]))$$

By classifying the content obtained within a specific time segment into entities of the SG graph, we generate a populated SG graph. This process is repeated for each time segment, yielding a sequence of knowledge graphs, each dedicated to a particular time segment (s_i). These graphs encapsulate the chronological evolution of an individual's behavior, emotions, or environmental perception, which in turn informs the identity inference module.

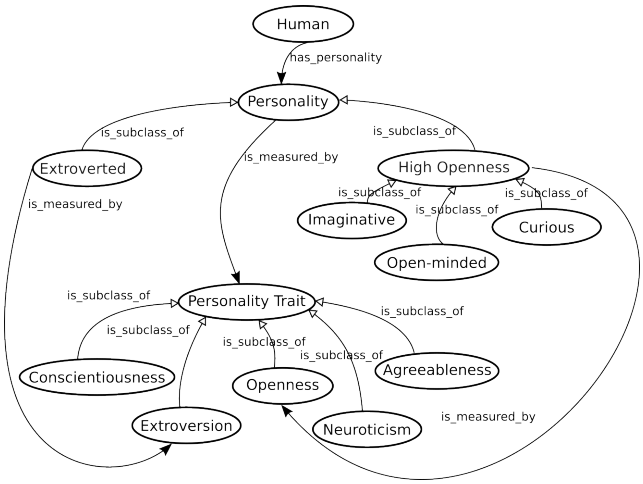

Figure 2: A view of the DUL ontology enriched with psychology's formal theories on identity and mental models

Identity Inference Module: The trained model produces a series of Situation Graphs (SGs), collectively representing a sequence of scenarios experienced by an individual. To fully leverage the potential of the structured information, we integrate the reasoning layer, consisting of the identity reasoning module, with an upper-level ontology known as DOLCE Ultralite (DUL) (Gangemi et al. 2002). Specifically, DUL facilitates the representation of situations, which in our context correspond to the distinctly united sequential contents of SG graphs. The DUL ontology, as shown in Figure 2, is also enriched with formal theories from fields like psychology and cognitive science concerning concepts related to identity and mental models (Rahnama, Alirezaie, and Pentland 2021). These theories establish connections between various situations an individual may encounter (e.g., listening to a specific music genre while driving) and their corresponding personality traits (Ruth 2020). Below are simplified examples of facts and rules about the five main personality traits (*Openness*, *Conscientiousness*, *Extroversion*, *Agreeableness*, and *Neuroticism*) and their relations with different daily activities, extracted from psychological studies incorporated into the expanded DUL ontology:

O(x) : `x has a high degree of Openness to new exps.`
C(x) : `x has a high degree of Conscientiousness.`
E(x) : `x has a high degree of Extraversion.`
A(x) : `x has a high degree of Agreeableness.`
N(x) : `x has a high degree of Neuroticism.`

Situation(x, s) : `x encounters situation s.`

$\exists s\ [\text{Situation}(x, s) \wedge s(\text{creative\_activity}, \text{experimental\_music})] \Rightarrow \text{O(x)}$
$\exists s\ [\text{Situation}(x, s) \wedge s(\text{studying}, \text{instrumental\_music})] \Rightarrow \text{C(x)}$
$\exists s\ [\text{Situation}(x, s) \wedge s(\text{eco-activity}, \text{friends})] \Rightarrow \text{E(x)}$
$\exists s\ [\text{Situation}(x, s) \wedge s(\text{solitude}, \text{ambient\_sounds})] \Rightarrow \text{N(x)}$
$\exists s\ [\text{Situation}(x, s) \wedge s(\text{family\_gathering}, \text{food}, \text{entertainment})] \Rightarrow \text{A(x)}$

The seamless integration of rules and represented situations enables the model to effectively infer the most likely personality traits of an individual, drawing from their consistent behavioral patterns. The incorporation of additional rules and more detailed capture of situations enhances the accuracy of identity inference for the individual.

Decentralized Data & Distributed Learning:

The first challenge (**C1**) in constructing an individual's digital identity model is the task of efficiently collecting data from a multitude of sources while respecting user privacy, ensuring that personal information remains protected throughout. Our proposed model's architecture, as depicted in Figure 1, facilitates distributed data collection. This setup assumes that each channel and the subsequent embedding module are situated on separate devices in close proximity to the data. Importantly, our model is trained exclusively on an individual's digital footprint, with their consent.

Given the decentralized nature of digital platforms and the wide range of data types (e.g., text, images, videos, locations, environmental data, etc.) that require varied learning approaches, we advocate the adoption of federated (or collaborative) learning. This distributed machine learning technique, as discussed in (Lalitha 2018), allows decentralized platforms with localized data to collaboratively train a model without the need to share raw data.

Discussion

This paper proposes an approach for personalized models using structural learning, allowing people's expertise to become more transferable in the form of a knowledge lens or perspective. We believe approaches like these contribute to the formation of a new internet ecology in which users can use decentralized and more privacy-preserved capabilities to share knowledge with each other and within smaller trust circles, such as intergenerational families. We intend to continue disseminating our work with more data-driven experiments in subsequent publications and also highlight applications of this framework in areas such as education, media, augmented eternity, digital immortality, and politics.

Ethical Statement.

Our proposed idea of generating digital identity models carries both societal benefits, like enhanced data control and security, as well as potential risks, such as data security, data governance, consent, privacy and information sentiency. We are carefully and meticulously incorporating these risks into our research plans to carefully analyze, assess, validate and publish our approaches on how we overcome these challenges. We believe addressing these challenges will be another key contribution of our research in this and subsequent publications under our research agenda.

Acknowledgments

We would like to express our appreciation to the team at Flybits, Toronto Metropolitan University, The Creative School at TMU, and MIT Media Lab for their support of our ongoing research in this field.

References

Anders, M.; Obaidi, M.; Paech, B.; and Schneider, K. 2022. A Study on the Mental Models of Users Concerning Existing Software. In Gervasi, V.; and Vogelsang, A., eds., *Requirements Engineering: Foundation for Software Quality*, 235–250. Cham: Springer International Publishing. ISBN 978-3-030-98464-9.

Bochman, A. 2003. A Logic for Causal Reasoning. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, IJCAI'03, 141–146. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

d'Avila Garcez, A. S.; and Lamb, L. 2020. Neurosymbolic AI: the 3rd wave. *Artificial Intelligence Review*, 1–20.

Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Cite arxiv:1810.04805Comment: 13 pages.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; Uszkoreit, J.; and Houlsby, N. 2020. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale.

Fischer, G. 2001. User Modeling in Human–Computer Interaction. *User Modeling and User-Adapted Interaction*, 11: 65–86.

Gangemi, A.; Guarino, N.; Masolo, C.; Oltramari, A.; and Schneider, L. 2002. *Sweetening Ontologies with DOLCE*, 166–181. Berlin, Heidelberg: Springer. ISBN 978-3-540-45810-4.

Grover, A.; and Leskovec, J. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864.

Hamilton, W. L.; Ying, R.; and Leskovec, J. 2017. Inductive Representation Learning on Large Graphs. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, 1025–1035. Red Hook, NY, USA: Curran Associates Inc. ISBN 9781510860964.

Lalitha, A. 2018. Fully Decentralized Federated Learning.

Liu, W.; Zhou, P.; Zhao, Z.; Wang, Z.; Ju, Q.; Deng, H.; and Wang, P. 2020. K-BERT: Enabling Language Representation with Knowledge Graph. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(03): 2901–2908.

Nersessian, N. J. 1992. In the Theoretician's Laboratory: Thought Experimenting as Mental Modeling. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1992: 291–301.

Radford, A.; Kim, J. W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning Transferable Visual Models From Natural Language Supervision. In Meila, M.; and Zhang, T., eds., *Proceedings of the 38th International Conference on Machine Learning, ICML 2021, 18-24 July 2021, Virtual Event*, volume 139 of *Proceedings of Machine Learning Research*, 8748–8763. PMLR.

Rahnama, H.; Alirezaie, M.; and Pentland, A. 2021. A Neural-Symbolic Approach for User Mental Modeling: A Step Towards Building Exchangeable Identities. In *AAAI 2021 Spring Symposium on Combining Machine Learning and Knowledge Engineering, MAKE*.

Ruth, . M. D., N. 2020. Associations between musical preferences and personality in female secondary school students. *Psychomusicology: Music, Mind, and Brain*, 30(4): 202–211.

Treur, J.; and van Ments, L., eds. 2022. *Mental Models and their Dynamics, Adaptation, and Control: a Self-Modeling Network Modeling Approach*. Studies in Systems, Decision and Control (SSDC). Springer Nature Switzerland AG. ISBN 9783030858209.