Age Inference on both Textual and Social Perspectives with Semi-supervised Learning

Dan Wu*, **Zixuan Teng***, **Dong Zhang**†
dzhang@suda.edu.cn
School of Computer Science and Technology, Soochow University, Suzhou, China

Abstract

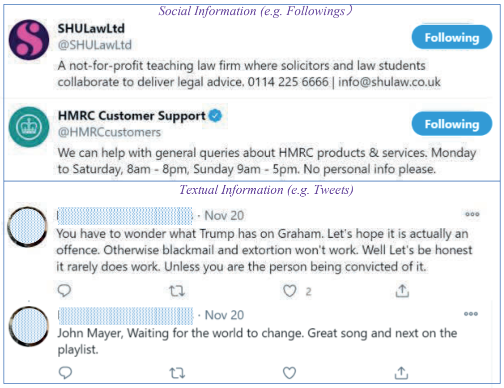
Large-scale annotated data is essential for age prediction in social media, yet obtaining such data is costly. Age is a key psychological and cognitive marker influencing communication and social behavior. Understanding age-related patterns in online interactions can provide insights into cognitive development and identity formation. To address data limitations, we propose a semi-supervised multimodal regression model leveraging Transformer-based variational autoencoders to infer age from textual and social features. This approach aligns with cognitive and social science theories on age-related behavioral patterns. Our framework effectively utilizes unlabeled data, reducing annotation dependency while enhancing predictive accuracy. Empirical results demonstrate superior performance over traditional classification and supervised baselines, advancing interdisciplinary research in age inference and online behavior modeling.

Keywords: Age inference; Cognitive modeling; Social behavior; Semi-supervised learning; Multimodal analysis.

Introduction

The popularity of social media platforms has significantly increased the demand for technologies capable of automatically analyzing users' profiles and their multimodal content (Chen, Cheng, Yang, et al., 2019). Age prediction, as one of the foundational tasks in this context, plays *a vital role in understanding the psychological and social aspects of online behavior*. In cognitive and social sciences, age is not merely a demographic feature but *a marker of developmental and cognitive stages* that influence how individuals express themselves and interact with others. For example, understanding age-related patterns in communication is crucial for cognitive theories that examine developmental changes in linguistic behavior and social engagement (Bleidorn et al., 2016; Suh, 2024) as shown in Figure 1. Moreover, in social contexts, age can significantly affect online behaviors such as posting frequency, topic preferences, and social interactions, which have broad implications for personalized marketing, user profiling, and social media analytics (Farnadi, Tang, Cock, & Moens, 2018; Zhu et al., 2024).

Traditionally, age prediction has been approached as a classification problem, dividing individuals into broad age groups (Liu, Nowson, & Perez, 2017). However, this simplistic categorization often overlooks the nuanced nature of age-related differences. More recently, scholars have shifted toward treating age as a continuous variable to facilitate finer-grained predictions that can better align with cognitive and psychological research that seeks to understand developmental differences in behavior over time (Flekova, Preotiuc-Pietro, & Ungar, 2016; Wang, Li, & Zhou, 2017; Krämer et al., 2024). The continuous nature of age, however, complicates the prediction task and requires models capable of making precise numerical estimates, rather than grouping users into predefined categories.



SHULawLtd
@SHULawLtd

A not-for-profit teaching law firm where solicitors and law students collaborate to deliver legal advice. 0114 225 6666 | info@shulaw.co.uk

HMRC Customer Support ✓
@HMRCcustomers

We can help with general queries about HMRC products & services. Monday to Saturday, 8am - 8pm, Sunday 9am - 5pm. No personal info please.

[redacted] · Nov 20
You have to wonder what Trump has on Graham. Let's hope it is actually an offence. Otherwise blackmail and extortion won't work. Well Let's be honest it rarely does work. Unless you are the person being convicted of it.

[redacted] · Nov 20
John Mayer, Waiting for the world to change. Great song and next on the playlist.

Figure 1: An example of user information in the social media.

One of the main challenges in age prediction, particularly when using social media data, is the scarcity of large-scale labeled datasets. Accurately annotating age from users' posts and social profiles is a time-consuming and costly process. Furthermore, some users may not post frequently or may only interact with others without contributing content themselves, making it difficult to infer their age from textual or social modality data alone (Nguyen et al., 2014; Zhang, Li, Wang, & Zhou, 2016; Fittipaldi et al., 2024). This issue is especially pertinent in psychological and cognitive studies, where fine-grained age prediction is necessary for understanding developmental stages that span a lifetime.

To address this challenge, we propose a semi-supervised learning approach to infer user age from both textual and social modalities. Semi-supervised learning is particularly suitable for scenarios where labeled data is limited, as it allows for leveraging large amounts of unlabeled data to improve prediction accuracy. In the context of social media, this

---

*Equal Contribution

†Corresponding Author: Dong Zhang

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

method has the potential to exploit the rich, unlabeled data that often exists in user profiles and posts to enhance model performance (Younis et al., 2023; Starck et al., 2024). Specifically, we introduce a semi-supervised multimodal regression model, the Multimodal VAE-R, which combines both textual and social features to predict age. Our model employs a variational autoencoder based on Transformer architectures, enabling it to learn from both labeled and unlabeled data while preserving the consistency between the two modalities. By using multimodal data, our approach aims to improve the precision of age prediction and address the limitations of unimodal approaches that only rely on one type of feature.

The contribution of this paper lies in the development of a semi-supervised multimodal regression model that effectively integrates textual and social features to infer age in social media contexts. Our empirical evaluations demonstrate that this approach outperforms several baseline methods in both regression accuracy and the ability to handle unlabeled data, *marking an important step forward in age inference for psychological and social applications*.

Related Work

Previous studies typically cast automatic age prediction as a classification problem, although this approach often suffers from ad hoc and dataset-dependent age boundaries (Cao et al., 2022). From a cognitive science perspective, age is not merely a numerical label but an essential factor influencing language use, emotional expression, and social behavior (Suh, 2024). Cognitive and developmental psychology suggest that as individuals age, their linguistic complexity, syntactic structures, and emotional expressiveness change systematically (Zhu et al., 2024). Younger individuals often use more expressive, informal language, while older individuals exhibit more structured, formal, and topic-driven communication styles (Han et al., 2023). These differences make age inference particularly relevant for studying cognitive and social development across the lifespan. Unlike traditional classification-based approaches, which categorize users into discrete age bins, regression-based models offer a more nuanced understanding of how age-related cognitive and social patterns manifest in digital communication. Recent works in psychology and computational linguistics emphasize that age prediction should be treated as a continuous variable to better capture age-related cognitive and behavioral shifts (Cameron et al., 2021). Therefore, we overview the related work on age inference by classification and age inference by regression, respectively.

Age Inference by Classification

Over the last decade, conventional studies on age prediction typically consider it as a classification problem and conduct their empirical research on two main domains, i.e., blog and social media (Chen, Cheng, Deng, Liang, & Qi, 2019; Chen, Cheng, Yang, et al., 2019). In blog domain, (Goswami, Sarkar, & Rustagi, 2009) and (Rosenthal & McKeown, 2011), explore both the textual and social modality information in automatic age classification. In social media domain, (Zhang et al., 2016) propose a supervised ensemble model to predict the age groups with multiple textual perspectives. More recently, (Weishampel, Staicu, & Rand, 2023) propose an interpretable, modeling framework to extract lower-dimensional relevant features of a user's posting behavior on social media, but still cast user attribute prediction as classification.

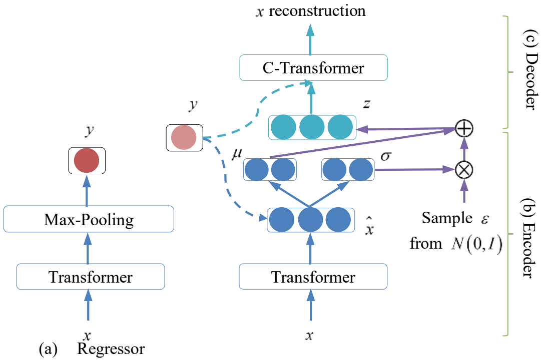


Figure 2: Unimodal VAE-R. Dashed line means the score of y is provided by the regressor when using unlabeled data, but when the labeled data is used, we employ the real observed score y.

Although age classification has been extensively explored in previous studies, related work on semi-supervised learning on age classification is rather few. (Ikeda, Takamura, & Okumura, 2008) propose a semi-supervised approach to age classification by training multiple sub-classifiers with textual features (Khare, Agrawal, Sharma, Awasthi, & Prakash, 2023).

Unlike their unimodal approaches, this paper proposes a semi-supervised multimodal approach. Moreover, this study focuses on a regression problem while their studies focus on the classification problem.

Age Inference by Regression

In contrast to age classification, much fewer studies model age prediction as a regression problem. (Flekova et al., 2016) present the study to investigate the correlations between the stylistic measures and exact age, in which age is viewed as continuous variables. From psychology and engineering perspectives (Bleidorn et al., 2016), finer grained prediction of user age is useful for downstream applications which use exact values (O'Connor et al., 2024).

To the best of our knowledge, no previous studies focus on semi-supervised regression for age inference. It is also worthwhile to note that, in the machine learning community, related studies on semi-supervised regression are also rare.

Methodology for Age Inference

In this section, we introduce our Unimodal VAE-R and Multimodal VAE-R approaches for age inference in details. Figure 2 and 3 shows our Unimodal VAR-R and Multimodal VAE-R approaches, respectively. Note that semi-supervised learning is particularly valuable in psychological and cognitive re-

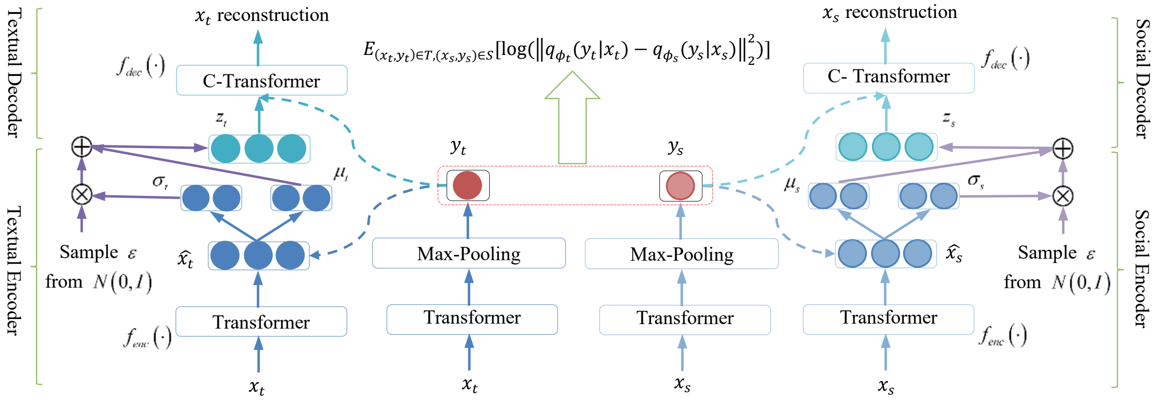


Figure 3: The overview of our proposed Multimodal VAE-R

search, where labeled datasets are often sparse, and behavioral patterns vary across cultural and social contexts (Chen, Stantic, & Chen, 2022). The ability to leverage large amounts of unlabeled social media data allows for more scalable and generalizable models that can adapt to individual differences in language processing, social cognition, and emotional regulation (Kim & Fingerman, 2022).

Prior research in cognitive psychology suggests that learning from limited labeled data is akin to human cognitive processes, where individuals generalize knowledge from partial observations (Suh, 2024). By incorporating semi-supervised learning into age prediction, we can model age-related cognitive and social traits in a way that mirrors natural learning mechanisms, making the approach not only technically robust but also cognitively plausible.

Unimodal VAE-R

Components. Basically, it consists of three main components: a regressor, an encoder network and a decoder network, which are corresponding to $q_\phi(y|x)$, $q_\phi(z|x,y)$ and $p_\theta(x|y,z)$. These networks are presented as follows.

(a) Regressor: For the regressor $q_\phi(y|x)$ as shown in Figure 2(a), we employ the Transformer network (Vaswani et al., 2017) to capture the semantic representation of user information and a max-pooling layer to obtain the user age representation for prediction, i.e.,

$$y = \mathrm{ReLU}(MP(Trans(x)))W^\top \tag{1}$$

where MP and $Trans(\cdot)$ denote the Transformer encoding and max-pooling, respectively. The output dimension of $MP(Trans(x))W^\top$ is 1, and $ReLU$ is the Rectified Linear Units.

(b) Encoder: For the encoder network depicted in Figure 2(b), each data pair (x,y) gets encoded into a soft ellipsoidal region within the latent space (Zhang, Li, Zhu, & Zhou, 2020). More precisely, the distribution of z is modeled as a diagonal Gaussian distribution $q_\phi(z|x,y)$:

$$\begin{aligned} \hat{x} &= f_{enc}(x), && (2)\\ q_\phi(z|x,y) &= N(\mu(\hat{x},y), diag(\sigma^2(\hat{x},y))), && (3)\\ z &\sim q_\phi(z|x,y) && (4) \end{aligned}$$

where $f_{enc}(\cdot)$ is the encoder function, implemented by $Trans(\cdot)$. $N(\cdot)$ denotes a Gaussian distribution with the reparameterization trick (Kingma, Mohamed, Rezende, & Welling, 2014).

(c) Decoder: As shown in Figure 2(c), the decoder is a conditional generative model (Zhang et al., 2020). It estimates the probability of generating x given latent variable z and real/predicted numbered value y:

$$p_\theta(x|y,z) = D(x|f_{dec}(y,z)) \tag{5}$$

where $f_{dec}(y,z)$ parameterizes a distribution D, usually a Gaussian distribution for input data. For implementing the decoder network, this paper adopts a potential conditional Transformer structure (W. Xu, Sun, Deng, & Tan, 2017). This structure integrates word embedding and score tensor at each time step, which is widely used (Serban, Sordoni, Bengio, Courville, & Pineau, 2016). We call this structure *C-Transformer*, which takes the concatenation of both words and regression score as input.

Loss Function. Through the joint training on both labeled and unlabeled samples, the loss is defined as follows,

$$\begin{aligned} \mathcal{J} &= \sum_{(x,y)\in C_l} \mathcal{L}(x,y) + \sum_{x\in C_u} \mathcal{U}(x) \\ &+ \alpha \mathbb{E}_{(x,y)\in C_l}[-\log q_\phi(y|x)] && (6) \end{aligned}$$

where C_l and C_u are labeled and unlabeled datasets respectively, α is a hyper-parameter of additional regression loss of labeled data. $\mathcal{L}(x,y)$ and $\mathcal{U}(x)$ are the variational bound of a single data point in the labeled data and the unlabeled data respectively (Zhang et al., 2020). Obviously, Unimodal VAE-R approach can be input with one of the textual modality and social modality, or even the features mixing of textual and social modalities.

Multimodal VAE-R

In order to make full use of the textual and social modality information, we can put the textual and social features together by concatenation with Unimodal VAE-R. However, textual and social modalities contain two different kinds of features and simply mixing them together does not seem a good way.

Instead, we propose a better way to perform semi-supervised regression with these two kinds of modalities by a novel model, namely *Multimodal VAE-R*. In this model, we first train two regressors with two kinds of features from textual and social modalities and then limit the two outputs from the regressors to be as similar as possible. The final objective function is defined as follows,

$$\begin{aligned}\mathcal{J} = &\sum_{(x_t,y)\in T_l} \mathcal{L}(x_t,y) + \sum_{x_t\in T_u} \mathcal{U}(x_t) \\ &+ \sum_{(x_s,y)\in S_l} \mathcal{L}(x_s,y) + \sum_{x_s\in S_u} \mathcal{U}(x_s) \\ &+\alpha\mathbb{E}_{(x_t,y)\in T_l}[-\log q_\phi(y|x_t)] \\ &+\beta\mathbb{E}_{(x_s,y)\in S_l}[-\log q_\phi(y|x_s)] \\ &+\gamma\mathbb{E}_{(x_t,y)\in T,(x_s,y)\in S}[\log(||q_\phi(y|x_t) - q_\phi(y|x_s)||_2^2)]\end{aligned} \quad (7)$$

where the symbols containing subscript t are related to the textual modality, and the symbols containing subscript s are related to the social modality.

Note that although we just leverage two modalities and limits the consistency of the output of these two modalities, in fact our model can be easily extended to three or more modalities by making the predicting of more than two modalities as the same as possible.

Experimentation

Experimental Settings

Data Settings: Our data is collected from Sina† randomly. We crawl each user's homepage which contains user information (e.g., *name, age, gender, followers, followings, verified type*), and their posted tweets (e.g., original and retweeted tweets). In total, we collect the homepages of about 5079 users, together with their posted messages and followings at a fixed period, such as June to December 2018. We randomly split the labelled, unlabeled, validation and test data as 5%/80%/5%/10%.

Feature Embedding (FE): To learn the feature representation, word content and word order of user-generated messages are used in the textual modality, while the following and follower IDs, and their descriptions are used in the social modality. We use LINE (Tang et al., 2015) to pre-train feature embedding based on all our crawled data with unsupervised learning.

Hyperparameters and Training: Here are the hyperparameters for training the model (based on validation-dataset loss). Word embedding dimension is 300. Both the encoder and decoder have a hidden size of 128. Latent variable z is 64 dimensional. Mini-batch size is 40. Hyper-parameters α, β, and γ range from 0.5 to 0.9. Models are trained end-to-end with the Adamax optimizer (Kingma & Ba, 2014), using a learning rate of 0.001.

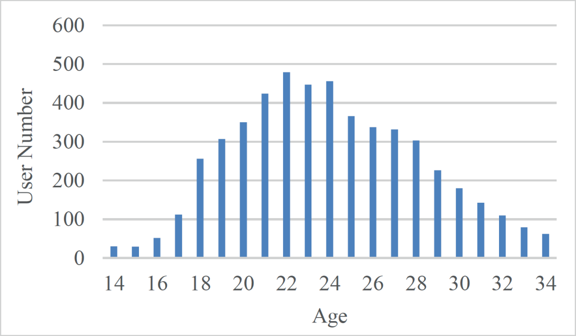


Figure 4: User distribution of different ages

Evaluation Metrics and Significance Test: Pearson's correlation coefficient (r, higher is better)(Nikolić, Mureşan, Feng, & Singer, 2012) and Mean Squared Error (MSE, lower is better)(Y. Li, Zha, & Zhou, 2017) between predictions and real ages are used as the performance metrics. Furthermore, t-test is used to evaluate the significance of the performance difference between two approaches.

Baselines and Implementation

Uni-modal Approaches. For thorough comparison, we implement the following approaches to user age inference when either textual or social modalities are employed. For all original classification approaches, in our implementation, the last layer is replaced with a fully-connected layer with non-linear activation function, similar to the Equation (1).

- **LSTM-R**: A supervised regression approach implemented with Long Short Memory Network for Regression (Graves et al., 2013), applied for textual and social perspectives respectively.
- **Self-SVR**: A semi-supervised learning approach which extends the supervised SVR approach based on self-training (Kostopoulos et al., 2018).
- **LP-TE**: The state-of-the-art of semi-supervised learning based on Transformer for either verbal or non-verbal input, which decomposes a neural network M into two components F and U. The layers in F are then frozen and only the layers in U will be updated during most time of the training (A. H. Li & Sethy, 2020).
- **GFDA:** The SOTA on social tasks formulated as regression problem (Weishampel et al., 2023).
- **GPT-4o mini:** A famous large language model (LLM) as SOTA on various social tasks (OpenAI, 2024).

Multi-modal Approaches. For thorough comparison, besides the unimodal approaches with early fusion, we implement the following approaches to age inference when both textual and social modalities are employed with late fusion or joint learning. For all original classification approaches, in our implementation, the last layer is replaced with a fully-connected layer with non-linear activation function, similar

---

[†] http://weibo.com/

Model	Textual		Social	
	$r\uparrow$	MSE $\downarrow$	$r\uparrow$	MSE $\downarrow$
LSTM-R (Graves, Jaitly, & Mohamed, 2013)	0.531	6.088	0.536	5.970
Self-SVR (Kostopoulos, Karlos, Kotsiantis, & Ragos, 2018)	0.480	6.370	0.466	6.963
LP-TE (A. H. Li & Sethy, 2020)	0.540	5.912	0.541	5.985
GFDA (Weishampel et al., 2023)	0.539	5.899	0.546	5.896
GPT-4o mini (OpenAI, 2024)	0.542	5.900	0.546	5.859
Unimodal VAE-R	**0.558**†	**5.601**†	**0.563**†	**5.675**†

Table 1: Experimental results of different regression approaches to user age inference with either textual or social modalities. The mark $\uparrow$ indicates that a larger value is better, while $\downarrow$ indicates that a lower value is better. The marker † refers to p-value < 0.05 when comparing with GFDA.

Model	$r\uparrow$	MSE$\downarrow$
LSTM-R (Textual+Social) (Graves et al., 2013)	0.548	5.806
UDMF (Farnadi et al., 2018)	0.561	5.769
CoREG (Z. Zhou & Li, 2007)	0.480	6.806
Weighted Voting (Z. H. Zhou, 2012)	0.501	5.964
SAFER (Y. Li et al., 2017)	0.572	5.636
LP-TE (Text+Social) (A. H. Li & Sethy, 2020)	0.561	5.552
MVSA (Q. A. Xu, Jayne, & Chang, 2024)	0.571	5.554
GPT-4o mini (Textual+Social) (OpenAI, 2024)	0.568	5.590
Unimodal VAE-R (Textual+Social)	0.555	5.728
Multimodal VAE-R	**0.596**†	**5.328**†

Table 2: Experimental results of different regression approaches to user age inference with both the textual and social modalities. The mark $\uparrow$ indicates that a larger value is better, while $\downarrow$ indicates that a lower value is better. The marker † refers to p-value < 0.05 when comparing with MVSA.

to the Equation (1).

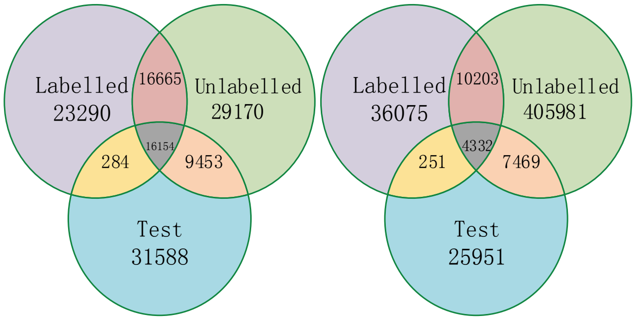
- **UDMF:** A supervised user attributes classification approach, considered as the state-of-the-art to age inference with multimodal learning(Farnadi et al., 2018).
- **CoREG:** A co-training style semi-supervised learning approach by (Z. Zhou & Li, 2007). This approach employs two $k-$nearest neighbor regressors based on two views respectively.
- **Weighted Voting:** An ensemble learning approach, which uniformly weights multiple regressors to perform semi-supervised regression (Z. H. Zhou, 2012). In our implementation, we build two regressors based on two modalities respectively.
- **SAFER:** A semi-supervised regression approach to learn a safe prediction from multiple semi-supervised regressors (Y. Li et al., 2017). This approach is considered as the state-of-the-art in semi-supervised regression. In our implementation, this approach is performed on the textual and social regressors.
- **MVSA:** The SOTA on multi-view learning for social tasks (Q. A. Xu et al., 2024).
- **GPT-4o mini (Textual+Socal):** We concatenate both the textual and social features as input fed into GPT-4o mini (OpenAI, 2024).
- **Unimodal VAE-R (Textual+Socal):** We concatenate both the textual and social features as input fed into our unimodal approach.

Results on Uni-modality

Table 1 shows the experimental results of different regression approaches to user age inference with either textual or social modalities. From this table, We can see that each approach using the features of textual modality achieves similar performance, compared to using social features. This result means that the two kinds of modalities are both helpful to predict user age. Therefore, the ability of just using one type of user information to predict age is relatively weak. Among all supervised learning approaches, **LSTM-R** performs best.

Additionally, we can observe that semi-supervised regression is challenging since the **self-SVR(TF-IDF)** performs much worse than **LSTM-R** and **LP-IE** performs not better than **LSTM-R**. As large language models (LLMs), **GPT-4o mini** performs better than most baselines on r, while a little worse on MSE. This is mainly due to the small number of labeled data, which are not sufficient for it to conduct in-context learning.

In contrast, the **Unimodal VAE-R** approach outperforms all the other six approaches, which demonstrate that our semi-supervised approach could leverage the unlabeled data to improve regression performance. Significance test shows that



(a) Textual (b) Social

Figure 5: The statistics of textual and social features' number in labelled, unlabelled and test sets respectively

the improvement of our unimodal semi-supervised approach over **LP-TE** and **LSTM-R** is significant (p-$value < 0.05$).

Results on Multi-modality

Table 2 shows the experimental results of different regression approaches to user age inference with both the textual and social modalities. From this table, we can see that, compared to using only one type of modality, using both textual and social modalities always achieves better performance. For instance, the **LSTM-R (Text+Social)** approach outperforms unimodal *LSTM-R* with a large margin. In the supervised baselines, **UDMF** with multimodal learning performs best. This indicates that multimodal features need to well capture the uni-modal independent and multimodal dependent characteristics. In the semi-supervised baselines, the **CoREG** approach fails to improve the performance with two kinds of features and it is even worse compared to the baseline **Self-SVR**. This may be due to the fact that co-training normally brings in the error transition. While **SAFER** outperforms other baselines except on MSE compared with **LP-TE** and MVSA, suggesting that: First, **SAFER** is effective to build a regressor on each modality; Second, **LP-TE** based on Transformer and MVSA based on multi-view learning also have certain advantages in age inference. For LLMs **GPT-4o mini**, it outperforms most baselines on metric r, but performs a litter worse on MSE. This suggest that this method is unstable and there is an urgent need to design a comprehensive approach.

Among all approaches, our approach, **Multimodal VAE-R**, performs best. Particularly, it outperforms the best-performed supervised baseline **UDMF** by 3.5% and 0.441 in terms of r and MSE respectively. Besides, it outperforms the best-performed semi-supervised baseline **MVSA** by 2.5% on r and **LP-TE** by 0.226 on MSE. Significance test shows that the improvement of our **Multimodal VAE-R** approach over **Unimodal VAE-R** and other strong baselines is significant (p-$value < 0.05$).

Further Analysis

From Figure 5, it can be seen that a large number of the features in the test samples do not appear in the labelled samples, but there are many overlapping features with the unlabelled samples, in either textual or social modalities. In other words, when we use unlabelled information to help age inference, it is more likely to improve the prediction performance. According to the previous subsection, our experimental results are consistent with this hypothesis.

Discussion

From a cognitive science perspective, age is a critical marker of developmental and psychological stages, influencing how individuals express themselves and interact with others. Our model captures these age-related patterns by leveraging linguistic and social features, which are known to evolve systematically with age (Cameron et al., 2021). For example, younger individuals tend to use more informal, expressive language, while older individuals exhibit more structured and topic-driven communication styles. These findings are consistent with cognitive theories that emphasize the role of age in shaping linguistic complexity, emotional expression, and social engagement.

Our semi-supervised learning framework also mirrors human cognitive processes, where individuals generalize knowledge from partial observations. By incorporating unlabeled data, our model mimics the way humans learn from limited information, making it not only computationally effective but also cognitively plausible. This alignment with natural learning mechanisms enhances the model's applicability to psychological and cognitive research, where understanding developmental trajectories is crucial (Cao et al., 2022).

A practical example of our model's utility can be seen in educational settings. Imagine a scenario where an online learning platform aims to tailor its content to different age groups. By using our Multimodal VAE-R model, the platform can infer the age of its users based on their posts and interactions, allowing it to deliver age-appropriate content. For instance, *younger students (e.g., Grade 2 normally corresponds to 8 years old) might benefit from more interactive and visually engaging materials*, while *older learners might prefer structured and detailed resources (e.g., Grade 10 normally corresponds to 16 years old)*. This personalized approach can enhance learning outcomes by aligning content with the cognitive and developmental needs of users.

Conclusion

Our work not only advances the technical capabilities of age prediction in social media but also contributes to the broader field of cognitive science by providing a computational tool to study age-related cognitive and social behaviors. By bridging computational methods with psychological theories, we offer a framework that is both technically robust and cognitively meaningful, paving the way for future research in understanding human development and behavior in digital environments. Additionally, investigating the cultural and contextual factors that influence age-related online behaviors could also provide deeper insights into cognitive and social development across diverse populations.

References

Bleidorn, W., Arslan, R. C., Denissen, J. J., Rentfrow, P. J., Gebauer, J. E., Potter, J., & Gosling, S. D. (2016). Age and gender differences in self-esteem—a cross-cultural window. *Journal of personality and social psychology*, *111*(3), 396.

Cameron, D., de Saille, S., Collins, E. C., Aitken, J. M., Cheung, H., Chua, A., … Law, J. (2021). The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. *Computers in human behavior*, *114*, 106561.

Cao, Y., Berend, D., Tolmach, P., Amit, G., Levy, M., Liu, Y., … Elovici, Y. (2022). Fair and accurate age prediction using distribution aware data curation and augmentation. In *Proceedings of the ieee/cvf winter conference on applications of computer vision* (pp. 3551–3561).

Chen, J., Cheng, L., Deng, B., Liang, J., & Qi, J. (2019). Cross-media age regression with textual adaptation. *Journal of Physics Conference Series*, *1168*, 032019.

Chen, J., Cheng, L., Yang, X., Liang, J., Quan, B., & Li, S. (2019). Joint learning with both classification and regression models for age prediction. *Journal of Physics: Conference Series*.

Chen, J., Stantic, B., & Chen, J. (2022). Age prediction of social media users: case study on robots in hospitality. In *International conference on robot intelligence technology and applications* (pp. 426–437).

Farnadi, G., Tang, J., Cock, M. D., & Moens, M. (2018). User profiling through deep multimodal fusion. In *Proceedings of WSDM 2018* (pp. 171–179). Retrieved from `https://doi.org/10.1145/3159652.3159691` doi: 10.1145/3159652.3159691

Fittipaldi, S., Legaz, A., Maito, M., Hernandez, H., Altschuler, F., Canziani, V., … others (2024). Heterogeneous factors influence social cognition across diverse settings in brain health and age-related diseases. *Nature Mental Health*, *2*(1), 63–75.

Flekova, L., Preotiuc-Pietro, D., & Ungar, L. H. (2016). Exploring stylistic variation with age and income on twitter. In *Proceedings of ACL 2016.* Retrieved from `http://aclweb.org/anthology/P/P16/P16-2051.pdf`

Goswami, S., Sarkar, S., & Rustagi, M. (2009). Stylometric analysis of bloggers' age and gender. In *Proceedings of ICWSM 2009.* Retrieved from `http://aaai.org/ocs/index.php/ICWSM/09/paper/view/208`

Graves, A., Jaitly, N., & Mohamed, A.-r. (2013). Hybrid speech recognition with deep bidirectional lstm. In *2013 ieee workshop on automatic speech recognition and understanding* (pp. 273–278).

Han, N., Li, S., Huang, F., Wen, Y., Wang, X., Liu, X., … Zhu, T. (2023). Sensing psychological well-being using social media language: Prediction model development study. *Journal of Medical Internet Research*, *25*, e41823.

Ikeda, D., Takamura, H., & Okumura, M. (2008). Semi-supervised learning for blog classification. In *Proceedings of AAAI 2008* (pp. 1156–1161). Retrieved from `http://www.aaai.org/Library/AAAI/2008/aaai08-183.php`

Khare, O., Agrawal, K., Sharma, A., Awasthi, S., & Prakash, A. (2023). Age prediction on social media using machine learning and ai. In *2023 5th international conference on advances in computing, communication control and networking (icac3n)* (pp. 486–492).

Kim, Y. K., & Fingerman, K. L. (2022). Daily social media use, social ties, and emotional well-being in later life. *Journal of social and personal relationships*, *39*(6), 1794–1813.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, *abs/1412.6980*. Retrieved from `http://arxiv.org/abs/1412.6980`

Kingma, D. P., Mohamed, S., Rezende, D. J., & Welling, M. (2014). Semi-supervised learning with deep generative models. In *Proceedings of nips 2014* (pp. 3581–3589). Retrieved from `http://papers.nips.cc/paper/5352-semi-supervised-learning-with-deep-generative-models`

Kostopoulos, G., Karlos, S., Kotsiantis, S., & Ragos, O. (2018, 06). Semi-supervised regression: A recent review. *Journal of Intelligent & Fuzzy Systems*, *35*, 1-18. doi: 10.3233/JIFS-169689

Krämer, C., Stumme, J., da Costa Campos, L., Dellani, P., Rubbert, C., Caspers, J., … Jockwitz, C. (2024). Prediction of cognitive performance differences in older age from multimodal neuroimaging data. *Geroscience*, *46*(1), 283–308.

Li, A. H., & Sethy, A. (2020). Semi-supervised learning for text classification by layer partitioning. In *ICASSP 2020* (pp. 6164–6168). Retrieved from `https://doi.org/10.1109/ICASSP40776.2020.9053565` doi: 10.1109/ICASSP40776.2020.9053565

Li, Y., Zha, H., & Zhou, Z. (2017). Learning safe prediction for semi-supervised regression. In *Proceedings of AAAI 2017* (pp. 2217–2223). Retrieved from `http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14587`

Liu, F., Nowson, S., & Perez, J. (2017). A language-independent and compositional model for personality trait recognition from short texts. In *Proceedings of EACL 2017* (pp. 754–764). Retrieved from `https://aclanthology.info/papers/E17-1071/e17-1071`

Nguyen, D., Trieschnigg, D., Doğruöz, A. S., Gravel, R., Theune, M., Meder, T., & de Jong, F. (2014). Why gender and age prediction from tweets is hard: Lessons from a crowdsourcing experiment. In *Proceedings of COLING 2014* (pp. 1950–1961). Retrieved from `http://aclweb.org/anthology/C/C14/C14-1184.pdf`

Nikolić, D., Mureşan, R. C., Feng, W., & Singer, W. (2012). Scaled correlation analysis: a better way to compute a cross-correlogram. *European Journal of Neuroscience*, *35*(5), 742–762.

O'Connor, K., Golder, S., Weissenbacher, D., Klein, A. Z., Magge, A., & Gonzalez-Hernandez, G. (2024). Methods and annotated data sets used to predict the gender and age of twitter users: Scoping review. *Journal of Medical Internet Research*, *26*, e47923.

OpenAI. (2024). Gpt-4o mini: advancing cost-efficient intelligence. *https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/*.

Rosenthal, S., & McKeown, K. (2011). Age prediction in blogs: A study of style, content, and online behavior in pre- and post-social media generations. In *Proceedings of acl 2011* (pp. 763–772). Retrieved from `http://www.aclweb.org/anthology/P11-1077`

Serban, I. V., Sordoni, A., Bengio, Y., Courville, A. C., & Pineau, J. (2016). Building end-to-end dialogue systems using generative hierarchical neural network models. In *Proceedings of aaai 2016* (pp. 3776–3784). Retrieved from `http://www.aaai.org/ocs/index.php/AAAI/AAAI16/paper/view/11957`

Starck, S., Kini, Y. V., Ritter, J. J., Braren, R., Rueckert, D., Mueller, T. T., et al. (2024). Atlas-based interpretable age prediction in whole-body mr images. *Machine Learning for Biomedical Imaging*, *2*(iMIMIC 2023 special issue), 2247–2267.

Suh, J. H. (2024). Multi-label prediction-based fuzzy age difference analysis for social profiling of anonymous social media. *Applied Sciences*, *14*(2), 790.

Tang, J., Qu, M., Wang, M., Zhang, M., Yan, J., & Mei, Q. (2015). LINE: large-scale information network embedding. In *Proceedings of WWW 2015* (pp. 1067–1077). Retrieved from `http://doi.acm.org/10.1145/2736277.2741093` doi: 10.1145/2736277.2741093

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., … Polosukhin, I. (2017). Attention is all you need. In *Nips 2017* (pp. 5998–6008). Retrieved from `http://papers.nips.cc/paper/7181-attention-is-all-you-need`

Wang, J., Li, S., & Zhou, G. (2017). Joint learning on relevant user attributes in micro-blog. In *Proceedings of IJCAI 2017* (pp. 4130–4136). Retrieved from `https://doi.org/10.24963/ijcai.2017/577` doi: 10.24963/ijcai.2017/577

Weishampel, A., Staicu, A.-M., & Rand, W. (2023). Classification of social media users with generalized functional data analysis. *Computational Statistics & Data Analysis*, *179*, 107647.

Xu, Q. A., Jayne, C., & Chang, V. (2024). An emoji feature-incorporated multi-view deep learning for explainable sentiment classification of social media reviews. *Technological Forecasting and Social Change*, *202*, 123326.

Xu, W., Sun, H., Deng, C., & Tan, Y. (2017). Variational autoencoder for semi-supervised text classification. In *Proceedings of AAAI 2017* (pp. 3358–3364). Retrieved from `http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14299`

Younis, H. A., Ruhaiyem, N. I. R., Badr, A. A., Abdul-Hassan, A. K., Alfadli, I. M., Binjumah, W. M., … Nasser, M. (2023). Multimodal age and gender estimation for adaptive human-robot interaction: A systematic literature review. *Processes*, *11*(5), 1488.

Zhang, D., Li, S., Wang, H., & Zhou, G. (2016). User classification with multiple textual perspectives. In *Proceedings of COLING 2016* (pp. 2112–2121).

Zhang, D., Li, S., Zhu, Q., & Zhou, G. (2020). Multi-modal sentiment classification with independent and interactive knowledge via semi-supervised learning. *IEEE Access*, *8*, 22945–22954.

Zhou, Z., & Li, M. (2007). Semisupervised regression with cotraining-style algorithms. *IEEE Trans. Knowl. Data Eng.*, *19*(11), 1479–1493. Retrieved from `https://doi.org/10.1109/TKDE.2007.190644` doi: 10.1109/TKDE.2007.190644

Zhou, Z. H. (2012). *Ensemble methods: Foundations and algorithms.*

Zhu, X., Guo, C., Feng, H., Huang, Y., Feng, Y., Wang, X., & Wang, R. (2024). A review of key technologies for emotion analysis using multimodal information. *Cognitive Computation*, 1-27.