Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 193 (2021) 422–431

Procedia Computer Science

www.elsevier.com/locate/procedia

10th International Young Scientist Conference on Computational Science

MVAESynth: a unified framework for multimodal data generation, modality restoration, and controlled generation

Anton Lysenko[a,*], Irina Deeva[a], Egor Shikov[a]

[a]ITMO University, Saint-Petersburg, Russia

Abstract

Synthetic data generation is used nowadays in a number of applications with privacy issues, such as training and testing of systems for analyzing the behavior of social network users or bank customers. Very often, personal data is complex and describes different aspects of a person, some of which may be missing for some records, which makes it very hard to deal with. In this paper, we present MVAESynth, a novel framework for the data-driven generation of multimodal synthetic data. It contains our implementation of a multimodal variational auto-encoder (MVAE), which is capable of generating user multimodal personal profiles (for example, social media profiles data and transactional data) and training even with missing modalities. Extensive experimental studies of MVAESynth performance were conducted demonstrating its effectiveness compared with the available solutions for the following tasks 1) training on data with missing modalities; 2) generating realistic social network profiles; 3) restoring missing profile modalities; 4) generating profiles with the specified characteristics.

Keywords: Synthetic Data; Social Media; Transactional Data; Multimodal Variational Autoencoder; Predictive Model.

1. Introduction

The progress of research and development in the area of personal data analysis directly depends on the availability of data for modeling and analysis of human behavior. Existing datasets do not keep pace with the growing flows in human interaction systems like social media. There are some publicly available data fields like number of friends or public posts, but there are also confidentiality restrictions on the collection, exchange, or dissemination of new data from social networks like subscriptions or list of interests. It is also clear that there are serious confidentiality restrictions on the collection and analysis of transactional data. The creation of synthetic datasets is an urgent task — it can be used to come over privacy issues [23], for data enrichment when source data is scarce, for the creation of target groups data, also for human behavior simulators or generation of community data with certain characteristics.

* Anton Lysenko
E-mail address: blinkop@gmail.com

10.1016/j.procs.2021.10.044

Existing methods for generating personal data have long been used to model population dynamics [5, 14], transport models [2, 10], and also for medical modeling [30]. However, the direction related to human modeling in social media has only recently begun to develop [1, 26]. And here we are faced with the fact that often the generation models of such personal data deal with incomplete or different nature data (numerical, text, graphic). In this case, it makes sense to utilize deep learning models, such as GANs and autoencoders, that have been successfully used to generate images or texts [32, 7, 15].

As a case study, we consider the task of multiple modalities generation containing users' social network profiles, their interests' profiles, and transactional data. Here, a modality is defined as a set of data obtained from a single source and describing a certain aspect of a person. The current example contains three modalities: personal data of the social network profile (number of friends, number of posts, number of videos on a page, etc.), the vector of interests, and transactional activity data provided by the partner bank. However, the approach is general since it imposes almost no requirements on the data. Studying the joint presentation of the modalities should give a more complete representation of the process or object being studied. We use a multimodal variational autoencoder (MVAE) that has a product-of-experts inference network and a sub-sampled training approach which makes it possible to flexibly mix different modalities and also to restore the missing modality based on other modalities.

2. Related Work

In recent years, many methods for generating synthetic personal data have been proposed in the literature. These methods can be divided into two categories corresponding to two different approaches: the first, called Synthetic Reconstruction (SR) [28], is aimed at creating a vector of characteristics for each agent in the synthetic population; the second, called Combinatorial Optimization (CO) [27] and consists in the duplication of known real individual records.

In the case of Synthetic Reconstruction, the joint distributions of the initial personal data attributes are evaluated. Algorithms such as IPF, IPU can be used for synthetic reconstruction [13]. Also, this approach can use structured probability models (Bayesian networks, Markov models, etc.) [25, 17].

Methods of Combinatorial Optimization are based on the selection of individuals from the sample with or without replacement, to satisfy the criteria of suitability. This convergence criterion is usually constructed using input features distribution. One can use genetic algorithms [18], greedy heuristics [24], annealing simulation algorithm [8], etc.

However, the mentioned methods for generating personal data are difficult to use when we have some missing data, as well as data of a different nature (numerical, textual, graphic). In this sense, deep learning models have great potential.

The works on the synthetic data generation with deep learning models mostly concern the generation of images (pictures, faces, etc.), [6] texts [20] and sounds [31] (for example, music). However, there are examples of reviews of deep learning applications in the medical field [22, 21]. At the same time, the problem of generating synthetic profiles of people, like social media profiles or psychological profiles, is poorly investigated. However, such a study could allow both the generation of artificial profiles for various services and algorithms and the interaction of various aspects of a person's personality. The difficulty is the generation of cross-modal profiles, in which data with different natures are aggregated (for example, bank data, posts and photos on social networks, demographic data). In [11], approaches based on autoencoders for the simultaneous generation of text and images, or the restoration of missing parts, are proposed. The problem is complicated by the fact that, unlike images, texts, and sounds, the quality of the generation of artificial profiles is difficult to evaluate for a person. Several metrics applicable to this problem were considered in [3]. However, some of these approaches require a discriminatory model to then measure the quality of its predictions on synthetic data.

3. Model

The task of generating synthetic user data can be formulated in terms of Bayesian statistics. Suppose we have x, the original vector of user data. Then the generative model has the form:

$$p_\theta(x, z) = p(z)p_\theta(x|z), \tag{1}$$

where z is the representation vector in a latent space. In formulation 1, $p(z)$ is a prior, usually spherical Gaussian, and $p_\theta(x|z)$ is a decoder, implemented as a deep neural network, with parameters θ, constructed with a simple likelihood (e.g. Bernoulli or Gaussian). The purpose of the training is to maximize the marginal probability of data ("evidence"). However, the task in such a formulation is difficult to process, the evidence lower bound (ELBO) is optimized instead, like it is done in variational autoencoders (VAE) [12, 16]. The *ELBO* is defined via an inference network, $q_\phi(z|x)$, which serves as a tractable importance distribution:

$$ELBO(x) \triangleq \mathbb{E}_{q_\phi(z|x)}[\lambda\ log(p_\theta(x|z))] - \beta\ KL[q_\phi(z|x), p(z)] \tag{2}$$

In formulation 2, KL$[p, q]$ is the Kullback-Leibler divergence between distributions p and q; β [9] and λ are weights balancing the parts in *ELBO*. *ELBO* is optimized using stochastic gradient descent with a reparameterization trick to backpropagate gradients through the random latent variable.

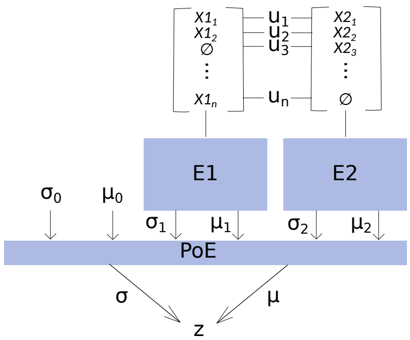


Fig. 1: MVAE architecture with 2 modalities. E_i, μ_i, σ_i — *i-th* inference network with it's variational parameters; μ_0 and σ_0 are the prior parameters. The product-of-experts (PoE) combines all variational parameters in the most efficient manner. For every user u_i the input data is given as a vector $X1_i$ and $X2_i$ or empty set sign if the modality is missing.

3.1. Multimodal variational auto-encoder

In the multimodal case we follow the approach of [29] and assume that $x_1, x_2, ..., x_N$ are the N conditionally independent modalities, given the common latent variable, z. We have a generative model of the form $p_\theta(x_1, x_2, ..., x_N, z) = p(z)p_\theta(x_1|z)p_\theta(x_2|z)...p_\theta(x_N|z)$. With such factorization, we can ignore the missing modalities when estimating the marginal probability. In Fig. 1, you can see a schematic representation of a multimodal architecture with two modalities.

Additional parameters μ_0 and σ_0 are the prior values for approximating joint-posterior in a product of experts (PoE). The PoE then combines μ and σ in the following way: $\mu = (a \sum_i \mu_i T_i)(\sum_i T_i)^{-1}$ and $\sigma = (\sum_i T_i)^{-1}$, where $T_i = V_i^{-1}$ is the inverse of i-th modality covariance. Inference networks are not restricted to have any fixed architecture and can vary for each modality. In our case, modalities are presented as one-dimensional vectors of social profile attributes, vector of interests, and the transaction profile, so the simple feed-forward neural networks were used.

Using the product of experts provides a possibility to learn the model in a case when some of the modalities are missing. As we learn the shared representation for each modality we suppose, that this representation is the same, if we encode the data of a particular user, and for that reason, the $ELBO$ loss function must be calculated through all of the missing modalities combinations. This solves the problem when we miss some of the linked data. Thus, we can train a model on missing some of the data in the way, presented in Fig. 1.

3.2. Factorized latent variable MVAE

Another model, that is based on multimodal variational autoencoder is factorized latent variable model [4] and in case of model name shrinking is called MEMVAE (Multi Embedding MVAE) during the current work. While utilizing only one latent space by MVAE, the MEMVAE forms two to each of the modality — modality-specific factors z_i and modality-invariant factors z_{shared} for *i-th* modality, which follows the intuition that each modality has its specific factors together with some factors, that are shared between the modalities. To infer z_i and z_{shared} the model utilizes one encoder for each modality but with concatenated output $\{z_i, z^{(i)}_{shared}\}$, then the PoE is used on a set of $z^{(i)}$ to produce the invariant representation z_{shared} — same as MVAE, while z_i are kept as is and holds he individual representations. The model is trained with the same approach as 3.1 with the only difference, that it has an additional D_{KL} term to each presented modality representation. To generate the *i-th* modality, the *i-th* decoder uses concatenated vector $\{z_i, z_{shared}\}$ and in case if a modality is missing z_i is filled with zeros.

4. Experiments and Results

The section provides the experiments to demonstrate, that the models implemented in the framework Section 3 are effective enough in solving the following tasks: 1) generation of a certain modality; 2) joint generation of multiple modalities; 3) recovery of a modality, given another; 4) generation with certain characteristics. Each subsection describes the experiments for one of the aforementioned aspects and shows the comparison of the MVAE models with the methods, which are applicable in a certain task.

4.1. Dataset description

For the experiments, a dataset was used, collected from the social network VK and with data from 80 thousand users. The data represent three modalities - socio-demographic data (age, gender, education, marital status, etc.), data of the user's interests, and the user's financial profile. The user's interests were collected by analyzing the groups to which he is subscribed, and represent the vector of the likelihood of interest in a particular topic; in total, 22 interests have been allocated for this vector (music, politics, beauty, etc.). The financial profile contains information about the average user spending for a certain period, the average amount spent per day, spending in certain categories, etc. All data is a table with discrete and continuous values. Since this is personal data, no links can be provided to it.

4.2. Generation of certain modality

While the log-likelihood seems to be the right choice as a metric for the generation ability, it's not that good because of the following circumstance: for a function of many variables, there are likely elements of the value space in which no events were observed. This can cause the metric to reach infinity. Therefore, it was decided to use the opposite approach, in which a metric is calculated that shows how confident the model is in a real test sample. For this, we evaluated the probability of some test sample to be generated with a given generative model and made a comparison between VAE and models from Section 3. To measure this probability we used evaluation metric from [29], where we calculate the probability as $log(p(x)) \approx log(\mathbb{E}_{q(z|x)}[\frac{p(x|z)p(z)}{q(z|x)}])$. This can be thought of as the inverse of a reconstruction error — the higher probability is the less reconstruction error. At the same time, we are interested in the variance of that probability, so both measures are presented in Table 1, where x_1, x_2 and x_3 are social, interests, and financial profiles respectively. These can be interpreted in the following way: a probability of observing a sample of any modality by MVAE and MEMVAE separately is pretty close to the VAE model, which was trained for both modalities separately.

Table 1: Probability of a test sample comparison.

Model	$log(p(x_1))$	$log(p(x_2))$	$log(p(x_3))$
VAE	-3.499	-5.299	-3.095
MVAE	-2.800	-5.297	-2.703
MEMVAE	-2.393	-2.189	-3.734
Model	$var(log(p(x_1)))$	$var(log(p(x_2)))$	$var(log(p(x_3)))$
VAE	5.458	4.414	7.572
MVAE	4.028	4.503	7.209
MEMVAE	2.951	2.530	13.736

Another way to compare the generation quality of a certain modality is to compare the distributions of the real data and generated one. The synthetic data was generated in the equivalent amount as the real (about 90k samples for each modality). Again, the comparison was made for the MVAE models and the VAE, which was learned for each modality separately. The results are shown for two fields for each modality — sex and number of followers from social media profile, the proportion of music and clothes interests in interest vector, parent indicator with the parent category amount from the transaction profile.

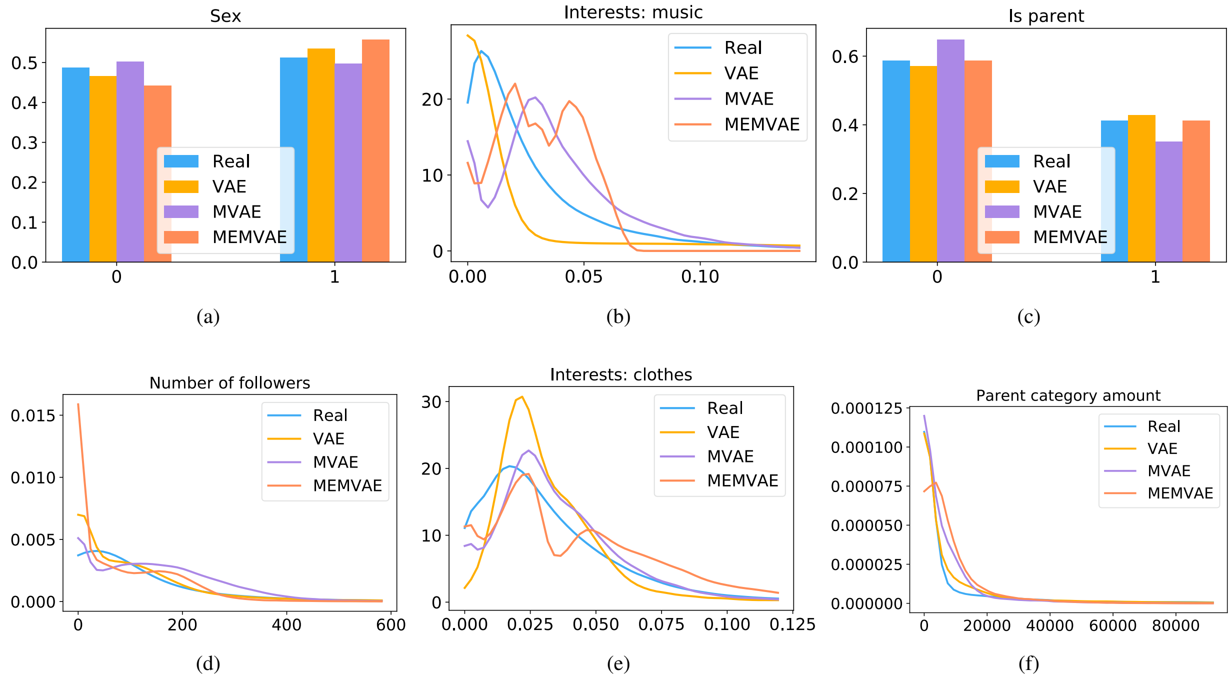


Fig. 2: Distributions comparison for some of the social media profile modality fields (2a, 2d), vector of interests modality fields (2b, 2e), transaction profile modality fields (2c, 2f).

4.3. Joint generation

When we generate multiple modalities from one latent vector we suppose that the generated modalities will reflect the same user. Two or more modalities in total might have the properties, which should be reproduced in the synthetic data and the example of such property is the pairwise correlation between the modalities. We will capture the linear correlation with the Kendall tau correlation coefficient and build the correlation matrices to watch if the correlations are observed in the synthetic data. The heat maps of the correlation matrices are presented in Fig. 3, where the value of matrix at position (i, j) shows the absolute value of correlation coefficient between *i-th* field of a first modality and *j-th* field of a second. Correlation between the real data modalities is compared to the correlation in synthetic data modalities, generated by MVAE and MEMVAE models. All the modalities were generated jointly from the same noise vector so that they must belong to one user. It is seen, that in both cases (MVAE and MEMVAE) some significant patterns are preserved, but the correlation values increase.

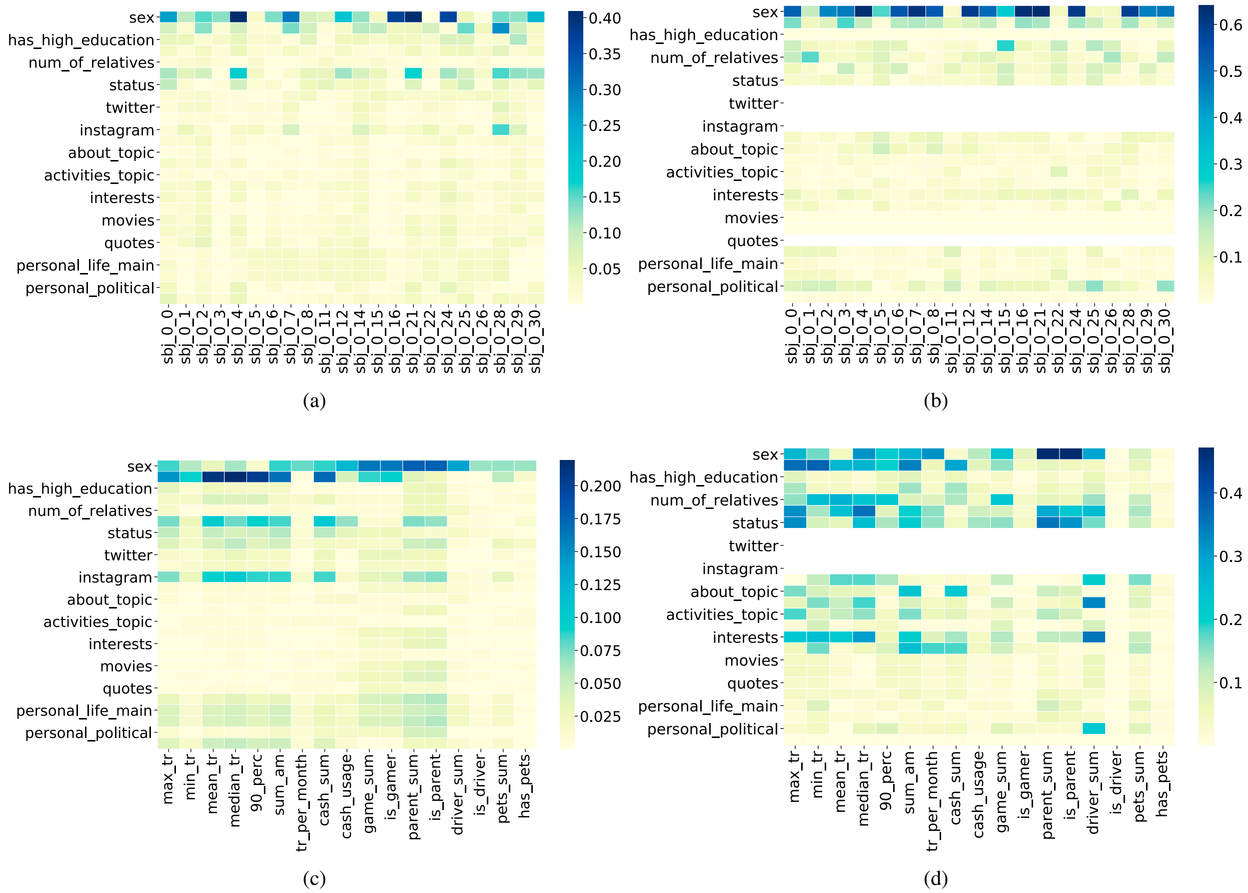


Fig. 3: Kendall tau correlation matrices between two modalities. 3a shows the correlation in real data for social media and interests profiles pair, while 3b shows the correlation for the same pair, but generated with MEMVAE. 3c and 3d represents the correlation between social media and transaction profiles pair for real and synthetic data, generated by MVAE, respectively.

While searching out more properties of the jointly presented modalities, we can let some models find those for us and show if they are presented in the synthetic data. To do this, a classifier was trained that distinguishes whether a given set of modalities belongs to a single user. The model for the classifier that was chosen is the gradient boosting method as it brings the highest ROC-AUC score among others. To train the classifier the real data was partly shuffled along with one or two modalities, and it was trained for every combination with two or more modalities. The results, obtained for each modality set case are presented in Fig. 4. Mostly, the ROC-AUC on the synthetic data, generated by the MVAE model is close to the score on the real data test set.

4.4. Data recovery

As it was described in Section 3, the MVAE and MEMVAE models can learn and infer the data with missing modalities. By that, we can consider these models as the modality recovery models, where some parts of the user profile are restored from others e.g. we can restore the social media profile given the user's transaction data and so on.

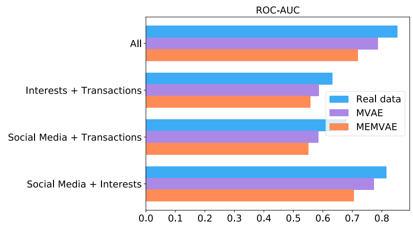


Fig. 4: ROC-AUC score for a classifier, that tells if given modalities belong to one user. For real data, the ROC-AUC was estimated on the test set.

To evaluate the quality of the data recovery, we simply can infer the model on some modality to generate the missing one and then compare the values of the generated modality with the values, predicted on the input modality by some model like gradient boosting, linear regression, etc. Specifically, we took each ordered pair of the modalities and predicted the one from another using either MVAE, MEMVAE, linear regression, random forest, or gradient boosting. Then we compared the resulting metrics on each of the fields: accuracy and ROC-AUC for the binary fields, accuracy for categorical fields, and the SMAPE and RMSE for continuous. Some of the comparison plots are illustrated in Fig. 5 — one for each of the metrics. In most cases, the results of MVAE are comparable, while MEMVAE on average gives the worst results when performing a recovery task.

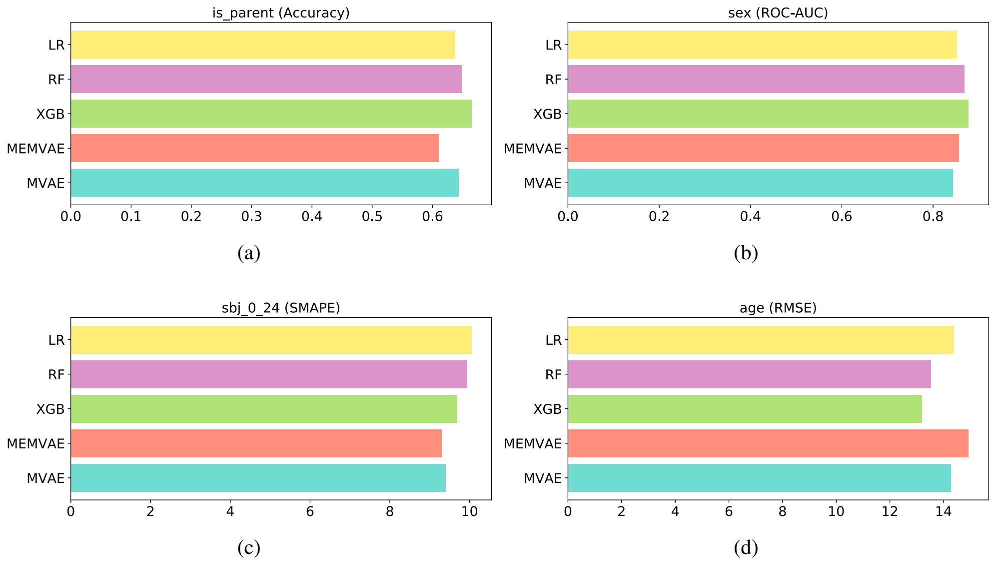


Fig. 5: Data recovery metrics comparison for some field of every modality. The multimodal autoencoder was compared to some classical machine learning models (LR - linear regression, RF - random forest, XGB - gradient boosting) in predicting the fields of a modality given another. The binary fields were compared with accuracy and ROC-AUC score while continuous with SMAPE and RMSE scores.

4.5. *Generation with certain characteristics*

To generate the profiles with certain characteristics we used the interests vector as a characteristics vector so that we can generate whether social media or transaction profile with the given fractions of interests at each point e.g. we can generate the social media profile of a user, who interested in business (where the interest in business is dominating), or transaction profile of a sportsman (domination of a sporting interest). Formally, we infer the model, giving only the interests modality and generate the social media and transaction modalities from the representation $p_\theta(x_1, x_3, z|x_2) = q_\phi(z|x_2)p_\theta(x_1|z)p_\theta(x_3|z)$, where x_1, x_2, x_3 are the social media, interests and transaction modalities respectively.

The experiment, which results are presented in Fig. 6, shows how a certain change in the vector of interests will affect some fields in remaining modalities. For that, we generated a random vector and increased or decreased the value of some point, related to the interest we want to manipulate. Then the vector was normalized to follow the format and was given to the model to generate the other modalities. The distributions of generated modalities with high/low level of certain interest were compared, where the pairs of a field and interest were taken according to the principle of correlation in real data e.g. we see the evidence of the clothes interests and sex correlation in the real data, so we want to keep this dependence and manipulate the proportion of men and women through the level of interest in clothes.

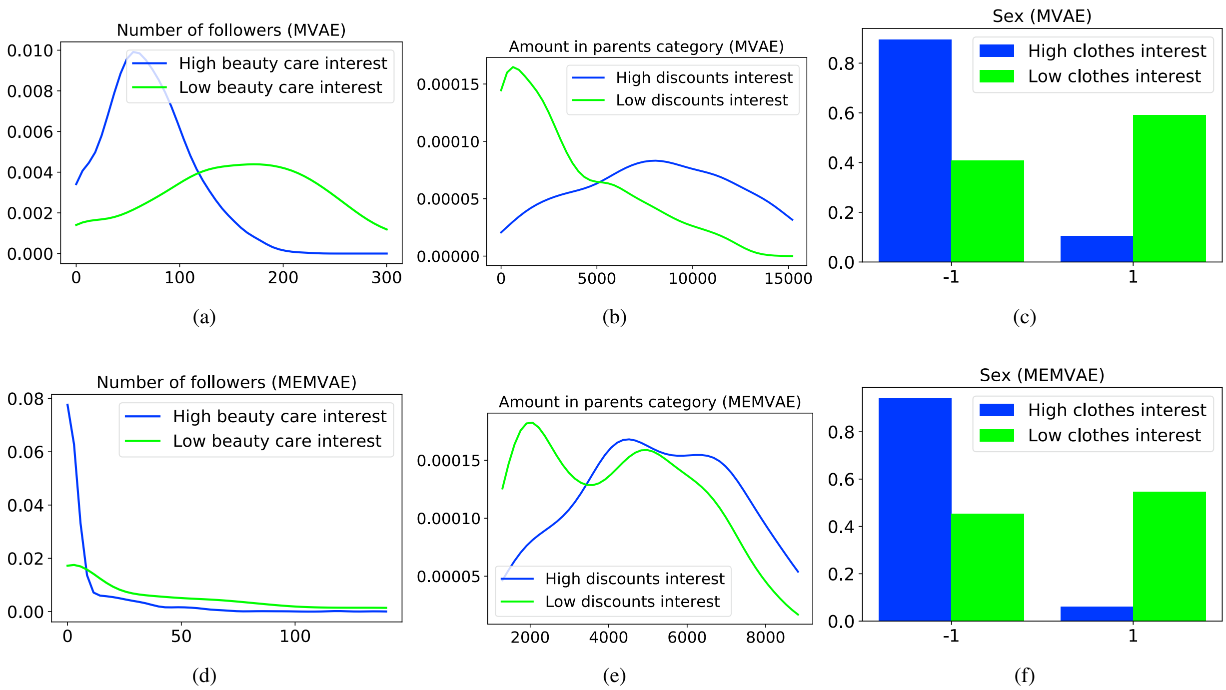


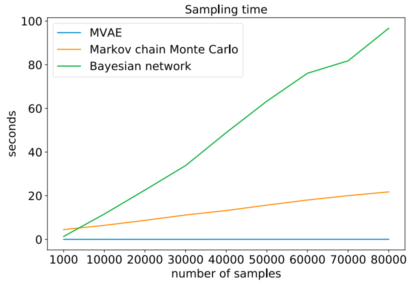
Fig. 6: Distribution comparison for some fields from social media and transaction profiles generated from interests vectors with a certain level of interest being increased/decreased.

4.6. Scalability

To proof the possibility of using the framework for large-scale data generation, an experiment was conducted to determine the generation time of a certain number of samples. The problem of generating a single modality of dimension 26 was considered. As baselines, we used well-known multidimensional generative approaches that are not based on neural networks: Bayesian networks and Markov chain Monte Carlo [19]. The results of this experiment are shown in Fig. 7. We can conclude that 1) the generation time depends linearly on the number of examples for all the methods; 2) our approach is several orders of magnitude faster than the presented presented baselines.

5. Framework structure

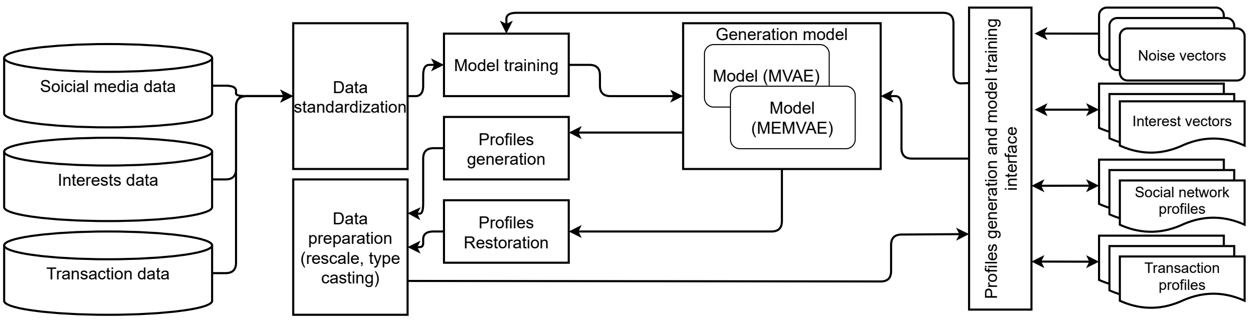
The multimodal synthetic data generation framework, that is built on top of the models from Section 3 solves the following tasks: 1) data generation from random noise vectors; 2) modalities recovery from the given ones; 3) generation of profiles with certain interests. It allows to choose what model to use: MVAE or MEMVAE (both are given already trained on the dataset from Section 4.1) and it also allows to train the model on the other dataset, which



(a)

Fig. 7: Sampling time comparison of the MVAE, Gibbs sampling, and Bayesian network models.

can be provided by the user. The schematic representation of a framework is illustrated in Fig. 8. The framework is open-sourced and available at Github [1].



(a)

Fig. 8: The MVAESynth framework scheme.

6. Conclusion

This paper introduces MVAESynth, which is a framework for the generation of multimodal synthetic data. Compared to available solutions, it supports different regimes (data generation, modality recovery, controlled generation). For all of these tasks, MVAESynth demonstrates competitive performance in terms of quality metrics. Also, scalability studies confirm that this approach allows you to generate thousands of examples in a fraction of a second. MVAESynth is open-sourced and available in the Github repository.

7. Acknowledgement

This research is financially supported by The Russian Science Foundation, Agreement №19-71-10078.

References

[1] Allen, T.T., Sui, Z., Parker, N.L., 2017. Timely decision analysis enabled by efficient social media modeling. Decision Analysis 14, 250–260.
[2] Anderson, P., Farooq, B., Efthymiou, D., Bierlaire, M., 2014. Associations generation in synthetic population for transportation applications: Graph-theoretic solution. Transportation Research Record 2429, 38–50.

[1] https://github.com/Blinkop/MVAESynth

[3] Borji, A., 2019. Pros and cons of gan evaluation measures. Computer Vision and Image Understanding 179, 41–65.
[4] Daunhawer, I., Sutter, T., Vogt, J.E., . Improving multimodal generative models with disentangled latent partitions .
[5] Geard, N., McCaw, J.M., Dorin, A., Korb, K.B., McVernon, J., 2013. Synthetic population dynamics: A model of household demography. Journal of Artificial Societies and Social Simulation 16, 8.
[6] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Advances in neural information processing systems 27.
[7] Han, C., Hayashi, H., Rundo, L., Araki, R., Shimoda, W., Muramatsu, S., Furukawa, Y., Mauri, G., Nakayama, H., 2018. Gan-based synthetic brain mr image generation, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), IEEE. pp. 734–738.
[8] Harland, K., Heppenstall, A., Smith, D., Birkin, M.H., 2012. Creating realistic synthetic populations at varying spatial scales: A comparative critique of population synthesis techniques. Journal of Artificial Societies and Social Simulation 15.
[9] Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A., 2017. beta-vae: Learning basic visual concepts with a constrained variational framework. Iclr 2, 6.
[10] Huynh, N., Namazi-Rad, M.R., Perez, P., Berryman, M., Chen, Q., Barthelemy, J., 2013. Generating a synthetic population in support of agent-based modeling of transportation in sydney .
[11] Karpathy A., F.F.L., 2015. Deep visual-semantic alignments for generating image descriptions. Proceedings of the IEEE conference on computer vision and pattern recognition , 3128–3137.
[12] Kingma, D.P., Welling, M., 2013. Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114 .
[13] Müller, K., Axhausen, K.W., 2010. Population synthesis for microsimulation: State of the art. Arbeitsberichte Verkehrs-und Raumplanung 638.
[14] Namazi-Rad, M.R., Huynh, N., Barthelemy, J., Perez, P., 2014. Synthetic population initialization and evolution-agent-based modelling of population aging and household transitions, in: International Conference on Principles and Practice of Multi-Agent Systems, Springer. pp. 182–189.
[15] Ranjan, A., Bolkart, T., Sanyal, S., Black, M.J., 2018. Generating 3d faces using convolutional mesh autoencoders, in: Proceedings of the European Conference on Computer Vision (ECCV), pp. 704–720.
[16] Rezende, D.J., Mohamed, S., Wierstra, D., 2014. Stochastic backpropagation and approximate inference in deep generative models. arXiv preprint arXiv:1401.4082 .
[17] Saadi, I., Mustafa, A., Teller, J., Farooq, B., Cools, M., 2016. Hidden markov model-based population synthesis. Transportation Research Part B: Methodological 90, 1–21.
[18] Said, L.B., Bouron, T., Drogoul, A., 2002. Agent-based interaction analysis of consumer behavior, in: Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1, pp. 184–190.
[19] Salvatier, J., Wiecki, T.V., Fonnesbeck, C., 2016. Probabilistic programming in python using pymc3. PeerJ Computer Science 2, e55.
[20] Semeniuta S., Severyn A., B.E., 2017. A hybrid convolutional variational autoencoder for text generation. arXiv preprint arXiv: 1702.02390. .
[21] Shamshirband, S., Fathi, M., Chronopoulos, A.T., Montieri, A., Palumbo, F., Pescapè, A., 2020a. Computational intelligence intrusion detection techniques in mobile cloud computing environments: Review, taxonomy, and open research issues. Journal of Information Security and Applications 55, 102582.
[22] Shamshirband, S., Fathi, M., Dehzangi, A., Chronopoulos, A.T., Alinejad-Rokny, H., 2020b. A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. Journal of Biomedical Informatics , 103627.
[23] Smith, M., Szongott, C., Henne, B., Von Voigt, G., 2012. Big data privacy issues in public social media, in: 2012 6th IEEE International Conference on Digital Ecosystems and Technologies (DEST), IEEE. pp. 1–6.
[24] Srinivasan, S., Ma, L., Yathindra, K., 2008. Procedure for forecasting household characteristics for input to travel-demand models. Technical Report.
[25] Sun, L., Erath, A., 2015. A bayesian network approach for population synthesis. Transportation Research Part C: Emerging Technologies 61, 49–62.
[26] Toriumi, F., Yamamoto, H., Okada, I., 2012. Why do people use social media? agent-based simulation and population dynamics analysis of the evolution of cooperation in social media, in: 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, IEEE. pp. 43–50.
[27] Williamson, P., Birkin, M., Rees, P.H., 1998. The estimation of population microdata by using data from small area statistics and samples of anonymised records. Environment and Planning A 30, 785–816.
[28] Wilson, A.G., Pownall, C.E., 1976. A new representation of the urban system for modelling and for the study of micro-level interdependence. Area , 246–254.
[29] Wu, M., Goodman, N., 2018. Multimodal generative models for scalable weakly-supervised learning, in: Advances in Neural Information Processing Systems, pp. 5575–5585.
[30] Xu, Z., Glass, K., Lau, C.L., Geard, N., Graves, P., Clements, A., 2017. A synthetic population for modelling the dynamics of infectious disease transmission in american samoa. Scientific reports 7, 1–9.
[31] Yang LC, Chou SY, Y.Y.M., 2017. A convolutional generative adversarial network for symbolic-domain music generation. arXiv preprint arXiv: 1703.10847. .
[32] Zhang, Y., Gan, Z., Carin, L., 2016. Generating text via adversarial training, in: NIPS workshop on Adversarial Training.