Energy profile representation in vector space

Christoforos Nalmpantis
Aristotle University of Thessaloniki
School of Informatics
Thessaloniki, Greece
christofn@csd.auth.gr

Odysseas Krystalakos
Aristotle University of Thessaloniki
School of Informatics
Thessaloniki, Greece
okrystal@csd.auth.gr

Dimitris Vrakas
Aristotle University of Thessaloniki
School of Informatics
Thessaloniki, Greece
dvrakas@csd.auth.gr

ABSTRACT

Word2vec is a computationally efficient way to represent words in vector space. Word embeddings are now considered an integral part of many Natural Language Processing problems. Its applicability has been successfully extended to other sequential problems, mainly in the discrete space. In this paper, the applicability of word2vec in time-series is examined, specifically for energy related problems. A novel framework, called Energy2vec, is presented and future improvements are discussed.

KEYWORDS

energy2vec, energy embeddings, word2vec, skip-gram

ACM Reference Format:
Christoforos Nalmpantis, Odysseas Krystalakos, and Dimitris Vrakas. 2018. Energy profile representation in vector space. In *SETN '18: 10th Hellenic Conference on Artificial Intelligence, July 9–15, 2018, Rio Patras, Greece.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3200947.3201050

1 INTRODUCTION

Neural embedding methods proved to be very efficient and boosted the performance in many linguistic tasks [10, 11]. These methods generate a geometric space, mapping the words to vectors and encapsulate semantic and syntactic relations. After achieving state of the art results, embedding methods have been adopted in other domains as well. Such domains include recommendation systems, translation, ranking of sets of entities, embedding graphs and others.

Barkan and Koenigstein described a method named Item2vec for recommendation systems based on Collaborative Filtering [2]. Item2vec uses skip-gram model with Negative Sampling and produces a vector space of items. Ozsoy et al. proposed non-textual features, namely the past check-ins of the users [13]. Bordes et al. proposed an energy-based model for learning vectors of entities, named TransE [4]. This system performs link prediction in knowledge bases and relationships of entities are represented as translations. Other embedding-based methods have also shown promising results in link prediction [5]. Wu et al. proposed a general-purpose model that can solve a wide variety of problems such as information retrieval, recommendation, link prediction in knowledge bases etc [14]. This model named StartSpace, showed state of the art or equivalent performance and is a general solution applicable to new problems.

The aforementioned applications have a common characteristic, they all use discrete sequential data. In this paper, we examine the applicability of Word2vec by using sequential data in continuous space. A novel framework is proposed, called Energy2vec, which describes how domestic energy data can be mapped to a vector space. The data come from the power consumption of domestic buildings and consequently the energy vector space is assumed to reflect some characteristics of the appliances' consumption. The framework is evaluated by analyzing the meaning of the produced vector space. In the end, suggestions are made on how to improve the current implementation and how to use it in order to solve energy related problems. Researchers are also encouraged to adopt the methodology to other time series domains.

2 RELATED WORK ON ENERGY DATA OF DOMESTIC BUILDINGS

The current proposal examines how to represent energy data in a vector space. Since the data come from domestic buildings, we examine the related work that has been done for the problem of power disaggregation. It is a fundamental problem in energy domain and is key to an efficient and accurate energy monitoring. Many solutions have been proposed to improve energy disaggregation and can be classified into two major methods: intrusive load monitoring (ILM) and non-intrusive load monitoring (NILM). A large-scale deployment favors NILM over ILM, because it offers lower costs, there is no need for multiple sensor configuration and installation is much simpler.

The modern approaches in NILM are based on Hidden Markov Models and most recently on Neural Networks [12]. Many of these approaches use non-traditional features such as contextual based, behavioral or indicators of the electric devices. In all cases the extra information to extract the features comes from external sensors, additional details of the building and the occupants or further computational analysis. Kim et al. proposed a conditional factorial hidden semi-Markov model, which is based on the observation that usage of appliance should have temporal patterns [8]. These patterns are based on the time of day and day of week. Wytock and Kolter proposed a contextual supervised source separation model [15]. The contextual features are radial basis functions over temperature and hour of day. Aiad and Lee suggested an unsupervised model which takes into account the interactions of devices [1]. Due to low quality of establishment, the operation of devices affects each

SETN '18, July 9–15, 2018, Rio Patras, Greece

ACM ISBN 978-1-4503-6433-1/18/07…$15.00
https://doi.org/10.1145/3200947.3201050

other, leading to disturbance of harmonics and interference currents. Integrating these interactions in a Factorial Hidden Markov Model improves substantially the disaggregation accuracy of the affected devices. Finally, Lange and Berges used binary subcomponents as features [9]. This is the first model that uses trained embeddings as features for the problem of power disaggregation. A neural decoder is used to extract the subcomponents, which are then combined to disaggregate the power consumption of the appliances.

Previous work on NILM shows that contextual features can improve the disaggregation results. Contextual features can vary from the occupants habits and weather conditions to power deficiencies and device interactions. Energy2vec tries to encapsulate the energy profile of a domestic building in a vector space, which potentially captures some contextual features.

3 GOALS OF THE PAPER

The goal of this paper is to propose a novel geometrical representation of household energy data, named Energy2vec. To the knowledge of the authors of this manuscript, no previous research exists modeling energy data in a geometric space.

Although Energy2vec is a domain specific framework, researchers are encouraged to adjust it to other non-textual and sequential data in continuous space.

4 DATA AND TOOLS

The energy data come from house 1 of UK-DALE dataset [7]. The dataset includes both the agreggated power consumption and the individual consumption of each appliance. The system was developed in Python, using NILMTK [3] for data processing. The model was developed with Tensorflow, in order to be able to deploy computation to one or more CPUs or GPUs. Finally, Tensorboard was used for illustration of diagrams, visualization of the vector space and evaluation of the model.

5 ENERGY2VEC

Energy2vec is based on Word2vec, with the difference that it is applicable on time series in continuous space. The neural network is trained using the Skip-gram [10] algorithm with Noise Contrastive Estimation (NCE) [6]. The framework is described by the following steps. Initially, the aggregated power signal has to be tokenized and the time series needs to be transformed to a discrete sequence of energy states, also called tokens. A collection of tokens is also built to consist the vocabulary of the final model. Next, the sequence of tokens can be used by the skip-gram model, which in turn formulates the embeddings. The result is a vector space. The framework is illustrated in Figure 1.

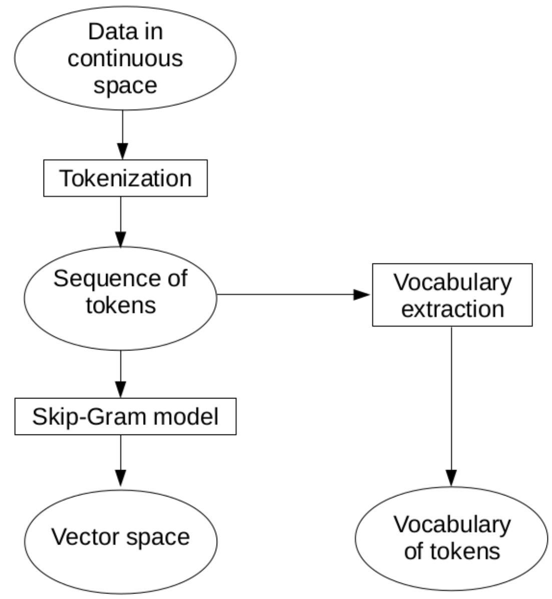

Figure 1: Energy2vec framework.

5.1 Tokenization

A discrete transformation of the energy data is necessary. A direct approach is to tokenize the data in a meaningful way. A natural supervised way of tokenization is to split the data into windows and describe the energy status of the house. The current implementation is restricted to a supervised labeling, because the disaggregated information is required. The length of each window is selected to be 1 minute and the energy description consists of the status of appliances during this time period. For simplicity, each device is considered to be at one of the following states: ON, START and STOP. START indicates the first time the device is switched on (after a period of being off). STOP indicates the first time the device is switched off (after a period of being on).

Next, tokens are created for the ON, START and STOP state for each device. For example for the appliance "kettle" the tokens would be: "kettle" when it is ON, "start_kettle" when it starts and "stop_kettle" when it stops. The total number of tokens will be three times the number of devices, and will consist our vocabulary. An additional token is called "noise", representing any energy noise. Additionally, the state when there is no energy consumption is called "zero state". The outcome of tokenization is actually a description of what is happening in the house during a specific time period. A tokenization example of a window is the following:

[start_kettle, fridge, stop_radio, oven]

That means that during this specific minute, someone started the kettle, the fridge was ON, the oven was ON and someone stopped the radio. After replacing a time series of energy data with a sequence of 1 minute tokenized windows, the continuous skip-gram model can be applied to project these tokens into a vector space.

5.2 Skip-gram

Building a sequence of meaningful tokens, results to a corpus, which describes the energy behavior of the appliances. Then, the skip-gram model is applied to learn vector representations of the tokens, based on the energy profile of the appliances. A vocabulary is defined by the collection of the tokens and a context is the window to the left and to the right of the target token. The goal of the algorithm is to predict the context, given the target. For example, the sequence [fridge, start_oven, radio, start_kettle, stop_television] with context 2 tokens would give the following pairs of target-context:

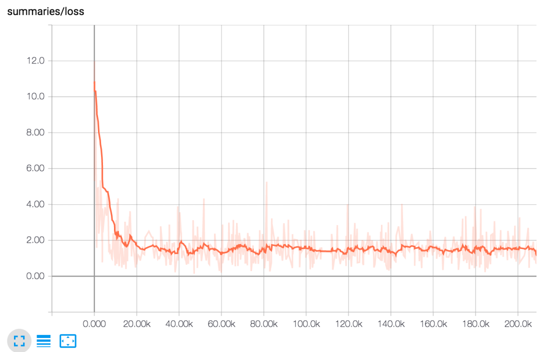

Figure 2: Loss function during training of Energy2vec.

(start_oven, [fridge, radio]), (radio, [start_oven, start_kettle]), (start_kettle, [radio, stop_television])

Let tkn_1, tkn_2,...tkn_T be a sequence of T training tokens. Then the objective function tries to maximize the average log probability according to the following formula:

$$1/T \sum_{t=1}^{T} \sum_{-c\leq_i \leq_c, i\neq_0} \log p(tkn_{t+i}|tkn_t) \qquad (1)$$

, where c is the training context.

As it is stated by Mikolov et al. [10], one limitation of the algorithm is that the order of the words doesn't affect the final result. However, this can be an advantage in the case of energy data, because the order is not important and there are no rules of syntax. For example, in a timeframe of a minute there is no difference if the token "fridge" is before or after the token "kettle".

The model is trained with Noise Contrastive Estimation (NCE) [6] and the optimizer is Adagrad with learning rate 0.01. The size of each embedding is 300 and the window is 6. A summary of the loss function during training on all the data of House 1 is shown in Figure 2.

From the loss function, it is noticed that the model learns very early the trained embeddings and then the learning rate becomes very slow. This can be attributed to the relatively low number of embeddings. A more granular tokenization would probably take longer to be trained.

6 RESULTS

It is showed that the popular Word2vec model is applicable on a time series and energy data can be represented in a continuous vector space. The intuition of this geometrical representation is the behavioral electricity consumption in a house. In the same way that Word2vec captures the meaning of words and their relation to each other, Energy2vec captures the appliance usage pattern.

For the evaluation, Tensorboard was used. It includes a similarity tool and a visualization tool, using PCA. Both Euclidean and cosine similarities were used and in most of the cases the results were very similar.

The validation process is the following. Firstly, the status of each device was plotted for all the training data in the period 2013-2016. The total time is almost 4 years. For a smaller period, the energy

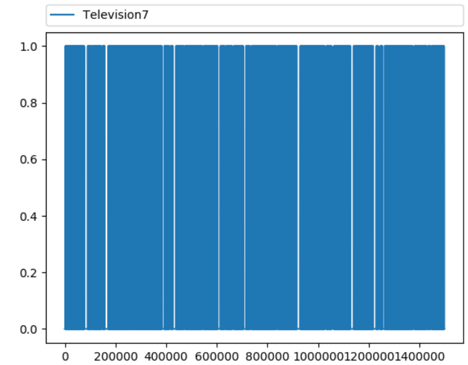

Figure 3

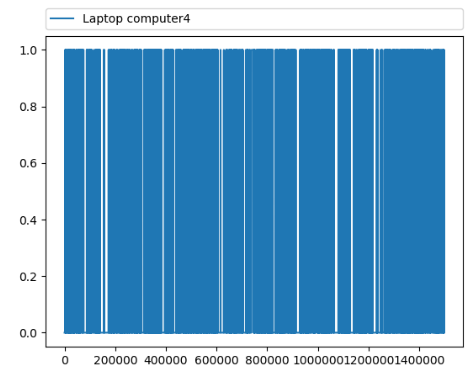

Figure 4

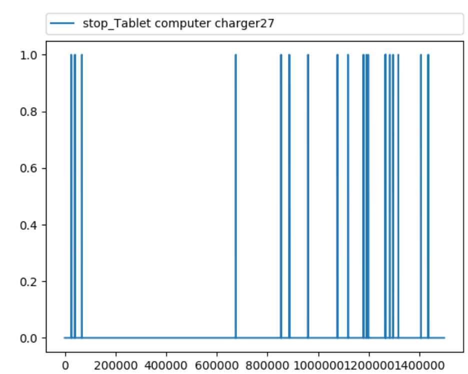

Figure 5

profile of the house would probably be different and vector space would have a different geometry. For each device, there are three different states, the working state, starting and stopping. Thus, there are three plots for each device, corresponding to the three learnt vectors named "appliance", "start_appliance" and "stop_appliance". The next step is to load the trained embeddings on Tensorboard and visualize them using PCA. Then, a vector can be selected and

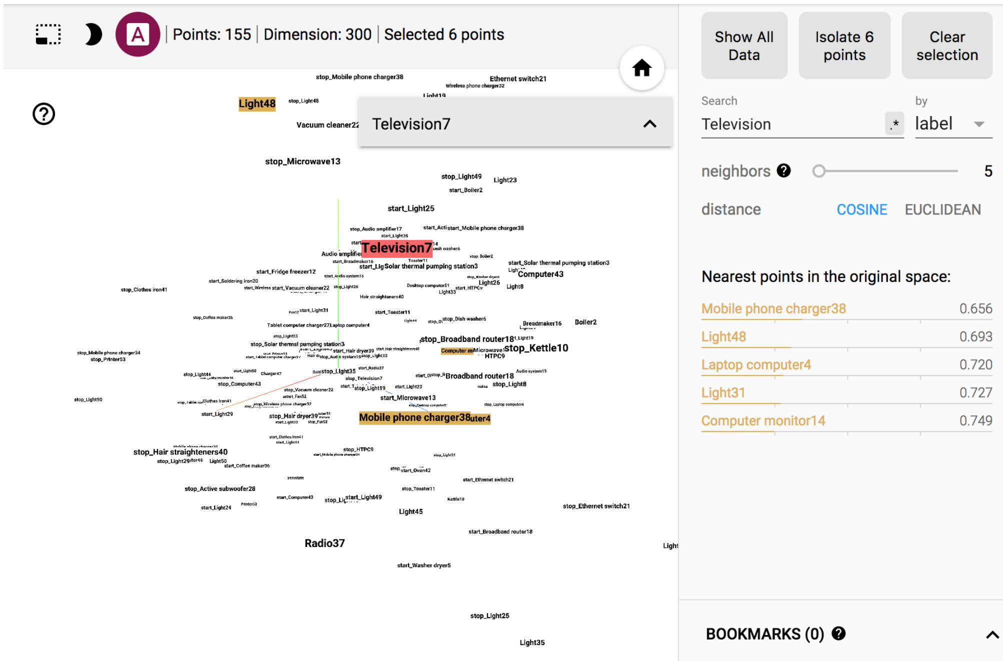

Figure 6

the top five similar vectors are found by means of Euclidean or cosine similarity. Finally, the respective plots of the top five similar vectors are compared. In the majority of the examples the results show that devices with similar plots are close in the geometrical space.

Figure 6 shows an example of cosine similarity search of the working state of television and Figure 3 shows the frequency diagram of television. The top similar energy state is presented in figure 4, depicting the frequency of usage of laptop computer. From these two images it can be observed that the residents were equally using both devices in a daily basis. In contrast, Figure 5 represents a device which was very rarely used and is much closer to the vector of zero state. It shows the frequency that tablet computer charger was unplugged and this could lead to the outcome that a tablet was rarely used.

Table 1 presents the results of four different similarity searches using Euclidean distance. The top five closest vectors are shown, for each of the following vectors: Television, Audio amplifier, Coffee maker and zero state. The same table shows the frequencies of each device state, which are also presented annually to show that there are different patterns in a different time period. Comparing the four groups of vectors it is obvious that vectors which are close in the geometrical space, also have similar total frequencies. This was expected as it is known that skip-gram is related to the frequencies of the tokens in a sequence. No pattern was identified regarding the change of the frequency of an energy state during time. This indicates that Skip-gram ignores the rate of change of frequency and seasonal energy behavior could not be captured.

From the experimental results there has been mainly identified three different groups of device states regarding frequencies: low frequency states, medium and high. This can also be seen from Table 1. The device states similar to zero state, belong to devices which are very rarely used or they are always OFF. The vectors that are similar to Television and Audio amplifier, represent devices that are used many times during a day or they are always ON. These include computers, fridge freezer, router etc. The vectors similar to Coffee maker belong to the medium frequency group. It is worth to mention that this group has vectors that could be characterized as high frequent. Such an example is Light58, which is close to both Coffee maker and Television. The model could possibly be improved, if there were more energy states and more data.

7 CONCLUSIONS

Energy2vec is the first attempt to transform energy profile of a building to a vector space. The vectors of this space depict some contextual information about the energy profile of the building. This information is directly related to the habits of the residents and the frequency they use each appliance. Another advantage of this model is that it is computationally efficient and reusable. Once an energy vector space is constructed, the trained embeddings can easily be used within a neural network. Thus, knowledge can be easily transfered to different models for various energy problems e.g. power disaggregation, prediction of energy consumption etc.

On the other hand, Energy2vec has some disadvantages. The vector space doesn't encapsulate any details about seasonal energy behavior or how the pattern of power consumption changes from time to time. Moreover, the tokenization step that was followed is a supervised process. In a real world environment labeled data are very rare and a scalable solution would require an unsupervised tokenization process. A possible solution, which is proposed for

Table 1: Similarity search results of energy states and their respective frequencies

Distance	Device States	2013	2014	2015	2016	Total
0	Television7	36260	52824	35963	25015	150062
1.265	Mobile phone charger38	103710	120430	289924	0	514064
1.286	Light48	12181	16100	8659	10037	46977
1.314	Laptop computer4	58081	72577	51753	31246	213657
1.334	Light31	98461	121358	57072	0	276891
1.344	Computer monitor14	43729	47400	24802	21072	137003
0	zerostate	0	0	0	0	0
1.284	Baby monitor46	0	0	0	0	0
1.296	start_Mobile phone charger34	0	0	0	0	0
1.299	stop_Tablet computer charger27	10	9	30	6	55
1.307	stop_Wireless phone charger32	19	81	0	3	103
1.309	start_Clothes iron41	21	7	4	4	36
0	Audio amplifier17	42389	84191	49104	37641	213325
1.297	Light25	65238	167924	140093	75298	448553
1.312	Solar thermal pumping station3	82008	103636	89678	37366	312688
1.313	start_solar thermal pumping station3	3002	3082	2410	808	9302
1.318	Broadband router18	0	0	0	137150	137150
1.328	start_Fridge freezer12	7445	9535	6534	3234	26748
0	Coffee maker36	17	25	975	968	3044
1.323	stop_Computer monitor14	457	746	719	453	2375
1.323	start_Ethernet switch21	3188	3049	1410	849	8496
1.326	start_Light48	101	106	53	63	323
1.339	Light48	12181	16100	8659	10037	46977
1.340	Ethernet switch21	17892	16460	6083	16008	56443

future work, would be to use a clustering algorithm in order to transform time series from continuous to discrete space.

From previous research it has been shown that embeddings can improve the performance of energy related problems, whereas in other domains such as natural language processing embeddings achieve state of the art results. Consequently, it is encouraged to use Energy2vec tackling problems in energy domain. Further research is advised to be conducted in order to improve the Energy2vec model and show experimental results in real world problems.

8 ACKNOWLEDGEMENTS

The authors would like to thank the Hellenic Artificial Intelligence Society (EETN) for its support for attending this conference.

This work has been funded by the ΕΣΠΑ (2014-2020) Erevno-Dimiourgo-Kainotomo 2018/EPAnEK Program 'Energy Controlling Voice Enabled Intelligent Smart Home Ecosystem', General Secretariat for Research and Technology, Ministry of Education, Research and Religious Affairs.

REFERENCES

[1] Misbah Aiad and Peng Hin Lee. 2016. Unsupervised approach for load disaggregation with devices interactions. *Energy and Buildings* 116 (2016), 96–103.

[2] Oren Barkan and Noam Koenigstein. 2016. Item2vec: neural item embedding for collaborative filtering. In *Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on*. IEEE, 1–6.

[3] Nipun Batra, Jack Kelly, Oliver Parson, Haimonti Dutta, William Knottenbelt, Alex Rogers, Amarjeet Singh, and Mani Srivastava. 2014. NILMTK: an open source toolkit for non-intrusive load monitoring. In *Proceedings of the 5th international conference on Future energy systems*. ACM, 265–276.

[4] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. 2013. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*. 2787–2795.

[5] Alberto Garcia-Duran, Antoine Bordes, and Nicolas Usunier. 2015. *Composing relationships with translations*. Ph.D. Dissertation. CNRS, Heudiasyc.

[6] Michael U Gutmann and Aapo Hyvärinen. 2012. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *Journal of Machine Learning Research* 13, Feb (2012), 307–361.

[7] Jack Kelly and William Knottenbelt. 2015. The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific data* 2 (2015), 150007.

[8] Hyungsul Kim, Manish Marwah, Martin Arlitt, Geoff Lyon, and Jiawei Han. 2011. Unsupervised disaggregation of low frequency power measurements. In *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 747–758.

[9] Henning Lange and Mario Bergés. 2016. The Neural Energy Decoder: Energy Disaggregation by Combining Binary Subcomponents. In *Proceeding of the 3rd International Workshop on Non-Intrusive Load Monitoring*.

[10] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[11] Andriy Mnih and Geoffrey E Hinton. 2009. A scalable hierarchical distributed language model. In *Advances in neural information processing systems*. 1081–1088.

[12] Christoforos Nalmpantis and Dimitris Vrakas. 2018. Machine learning approaches for non-intrusive load monitoring: from qualitative to quantitative comparation. *Artificial Intelligence Review* (2018), 1–27.

[13] Makbule Gulcin Ozsoy. 2016. From word embeddings to item recommendation. *arXiv preprint arXiv:1601.01356* (2016).

[14] Ledell Wu, Adam Fisch, Sumit Chopra, Keith Adams, Antoine Bordes, and Jason Weston. 2017. StarSpace: Embed All The Things! *arXiv preprint arXiv:1709.03856* (2017).

[15] Matt Wytock and J Zico Kolter. 2014. Contextually Supervised Source Separation with Application to Energy Disaggregation.. In *AAAI*. 486–492.