A research using NLP with a microscopic viewpoint to identify human activities for visually impaired people

Authors: Rubayet Mahjabin, rubayet.mahjabin@g.bracu.ac.bd, Nishat Zerin, nishat.zerin@g.bracu.ac.bd, Sameer Sadman Chowdhury, sameer.sadman.chowdhury@g.bracu.ac.bd, Md Sabbir Hossain, md.sabbir.hossain1@g.bracu.ac.bd, Md. Farhadul Islam, md.farhadul.islam@g.bracu.ac.bd, Annajiat Alim Rassel, annajiat@gmail.com

Abstract

Human activity recognition is currently an incredibly active research field in the development of advanced computing technology. In this project using HAR, the goal is to aid visually impaired individuals by allowing them to recognize human activities. Blind people may sense the people moving in front of them, but they can only sense a limited area. With HAR, their difficulty can be largely reduced, as it will allow them to recognize and comprehend the movements of human activity around them, which can help them when they are alone at home from some dangerous events. For example, robbery, or from other unfortunate events. This research attempts to apply the notion of natural language processing and the method of data symbolization to the study of human activity recognition. The studies are carried out creatively from a microscopic perspective, with an emphasis on the logic and relevance between data segments.

Introduction

This project aims to assist people who are blind or visually impaired in recognizing human activity by capturing a series of events that take place in front of them and delivering the output to them. HAR aims to recognize activities from a series of observations from the actions of humans and transmit that particular data to blind individuals by earphones. In this study, we talk about different ways to define activity elements and three algorithms. These include a dictionary-building approach, a corpus-building algorithm, and an activity recognition algorithm that uses a method for analyzing natural language called TFIDF. Using radio-frequency identification (RFID) technology, a system is used to figure out what a person is doing by figuring out how a signal changes.

Problem Statement

In this area, there are many models. Both traditional machine learning and the recently popularized neural networks have made major scientific advancements. The data mining algorithms used by conventional models for machine learning, like support vector machines (SVM) [2, 3], Bayesian models [4, 5], etc. Although neural networks have become a common machine learning paradigm in recent years, they are not without flaws. In real-world application contexts, its numerous calculations and manually set parameters provide challenging issues. Even at their best, traditional models still fall short of the performance of many neural network techniques, but they do strike a better equilibrium between the reliability of learning outcomes and the interpretability of learning models. Based on the aforementioned models, HAR has achieved numerous research successes. Setting an appropriate data segment length is a difficult task, which is a major issue. Most studies employ large segments because they can capture a whole movement's activity in one, hence they adopt a macroscopic viewpoint. For example, in the image below in Fig. 1 [6], all of these sub-figures together make a macroscopic walking movement.

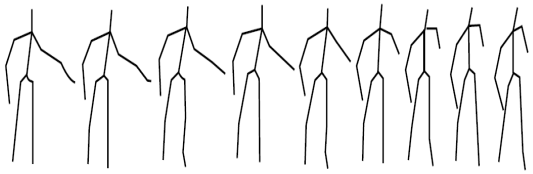

Fig. 1 A sequence of data segments with relevance and logic

To contain the entire activity for recognition from a macroscopic viewpoint, a suitable window length and data form must be selected; however, different activities may contain varying numbers of subfigures, and different activities' lengths may also vary. So, in this study, a microscopic solution is suggested. When activities are divided into appropriate little data segments, like the series of sub-figures in the image above, different groupings of data segments can then be combined to produce diverse activities. This point-of-view approach may provide a macroscopic solution to the problem. The primary objective of this study is to investigate HAR using appropriate data segments and combination forms from a microscopic perspective while addressing the limitations of a macroscopic perspective. The next sections of this essay will demonstrate how many comparison models that are based on separate data segments can better depict an activity than a single, isolated data segment.

Goal

In order to address this issue, the concept of natural language processing (NLP), currently the most prominent area, is used to analyze natural language data. A relevant and logical type of data is natural language. One benefit of NLP is that it may evaluate the consistency and logic of several data points before determining the semantics. The study of identifying human behavior in this paper can also use some NLP concepts and models. Activity-unit and activity-combination definitions for the symbols for activity-based data are offered. It describes how to convert a data segment into an activity unit. The method is based on a dictionary of activity units (combinations of activities) and the corpus of each activity. This approach corrects the flaw in single-segment recognition by determining the correspondence and logic between data units. The model in this study is built on the concept of data symbolization. Activity-TF-IDF, a better data analysis algorithm, is suggested in this research to fit the subject. In comparison to the original TF-IDF, this algorithm can improve recognition efficiency and better fit the study theme in this paper by calculating the correlation between the sample and corpus using word frequency. The microscopical TF-IDF activity recognition model (MiTAR) [1], a suggested human activity recognition model, incorporates the concept of natural language data analysis. It is a significant investigation into the HAR field.

Related Work

Traditional models, including SVM [2, 3], ELM [7], HMM [8], CRF [9], and Bayesian models [10], are widely employed in this discipline. The original SVM [2, 3] maintains a balance between performance and time cost in its role as a representative classical machine learning model. Before deep learning techniques are widely adopted, it might be the most often used model. It doesn't function well in many application contexts, though. Another typical model that is representative is the Hidden Markov Model (HMM) [8], but there are only a few dataset types for this model. Incorrect window length leads to incorrect markov chain length. Wrong probability results could be caused by an improper Markov chain length. The same restriction applies to Markov chain-like models like Bayesian models and CRFs (conditional random fields), among others.

Prior Knowledge

This section goes into great length on the background knowledge that went into this study, including some fundamental NLP definitions and the specifics of the original TF-IDF.

NLP fundamental definitions:

Alphabet: The most fundamental component of natural language is the alphabet.

Word: The letters make up the word. It is a semantically based data type. Even if a word can

convey meaning, a whole semantic entailment can be expressed in a single statement. A semantic component of natural language is the sentence.

Corpus: The most crucial component of NLP categorization is the corpus. It is a collection of phrases and words from a single type of language. Higher classification accuracy is guaranteed with a good corpus.

Dictionary: A dictionary is a list of terms that includes every word currently found in the corpus.

Maximum matching algorithm: This fundamental sentence segment method may divide a sentence into its component parts based on a dictionary.

TF-IDF

A language data analysis technique for NLP keyword extraction is called TF-IDF [11]. The primary meaning of an article is reflected in its key words. By determining the relationship between the keyword and the article, one can estimate the significance of a word in an article. In the discipline of NLP, a term is deemed significant for an article (corpus) and acceptable for classification if it occurs frequently in that article (corpus) but infrequently in other articles. This is the TF's guiding philosophy. The greater the IDF value, the lower the number of the article (corpus) containing the word is. This suggests that the word has a strong ability to differentiate between classes. This is the IDF's guiding principle. In this study, sample data and the corpus are compared after the corpus has been established in order to identify the sample data. Each activity's corpus is made out of a collection of words or sentences, while the activity unit and activity combination stand in for the letters of the alphabet and the word, respectively.

Microscopic TF-IDF activity recognition model (MiTAR)

This research uses an NLP-inspired technique to categorize and identify activity data. Each language in the field of NLP has its own corpus, and the current corpus can be examined using a variety of models to get the analysis findings.

Fundamental definitions of Microscopic TF-IDF activity recognition

In our categorization and recognition, the smallest unit is the **activity unit, which** reflects the original data segment's feature matrix from the M-window. In nature language, it is comparable to the "alphabet." M is the data segment's length.

A group of linked combined activity units is represented by an **activity combination**, which thus illustrates the relevance and logical connection between the data. It resembles "word" in the natural language.

A whole sequence created from the original sample data is called an **"activity sequence."**. In NLP, it has the same meaning as a phrase. The sample length is represented by the length of the activity sequence.

A text that was taken from the training data is represented by the **activity corpus**. One corpus is matched with one activity. The corpus is a group of words and word combinations from a single activity's training data. The calculation of the relationship between the corpus and the test sample forms the basis for activity recognition in this study.

The alphabet, word, sentence, corpus, and dictionary are the five core components of the model provided in this work since they are key components of natural language.The model described in our study can figure out human activity based on text (called a "corpus") and dictionaries.

Overview of MiTAR

These three primary algorithms are part of MiTAR: the activity recognition algorithm, the activity corpus building algorithm, and the activity dictionary building algorithm. Independent activity units are insufficient to precisely identify the activity, despite the fact that the original data has been translated into activity units. Individual activity units are weak and unconnected. As a result, using the first two algorithms, linked activity-combinations, vocabulary, and corpus are produced. The recognition algorithm then detects the activity label.

Algorithm

Steps of activity dictionary building algorithm [1]

- Transform all of the initial training data through data processing into activity-units.
- Assume that C is the word length. This shows that C is the number of activity units that can be included in a combination of activities. The letter C is a minor constant. This is necessary for the subsequent activities.
- By segmenting the collection of activity units produced from the original data according to word length, the activity combinations are produced.
- We treat two dissimilar activity combinations that have the same length and the same activity units as one activity combination, and if they contain the same activity units in one activity combination, they will be combined into one because the original data segment was brief and the order of the segments was random within a small range.
- The deduplicated activity combinations of the whole activity dictionary of the training data make up the final product. A dataset only has one dictionary. This dictionary contains all of the activity combinations in this dataset.

Steps of activity-corpus building algorithm[1]

- By using data processing, convert the initial training data into activity-units while maintaining the data's original order.
- Based on the range of training data for each activity, divide the unit sets of all the activities.
- Each activity's activity-units are transformed into activity combinations using the activity dictionary and maximum matching algorithms.
- Each activity's activity corpus is made up of activity combinations.
- Each action's activity corpus serves as a visual depiction of the result. Activity-corpus takes the shape of text.

Steps of activity recognition algorithm [1]

- Assume that the test sample is converted into a feature matrix and that the data segment window is initially set to M.
- In order to obtain appropriate activity-units, KNN is also used.
- The third step involves using the activity-dictionary with maximum matching method to segment the test sample activity-unit sequence.
- Next, using the Activity-TF-IDF, each corpus' correlation value with the test sample sequence is calculated.
- The label of the activity corpus with the highest correlation value is the identification result.

TF-IDF formula with improved compound word frequency [1]:

$$\text{Activity-TF} = \frac{N}{TNW}$$

$$\text{Activity-TF-IDF} = \log(\text{Activity-TF} + 1) * (\text{IDF}^2)$$

$$\text{FinalResult} = (\text{AllValue}) + WF * F$$

The number (**N)** denotes how many activity combinations of a particular movement there are in a corpus.
An experimental factor is **F**.
The total number of activity combinations in the sample across all corpora is referred to as **TNW**.
The value of Activity-TF-IDF for all activity combinations is referred to as **AllValue**.
WF is a representation of the ratio of all activity combinations in a corpus.

Activity Recognition with RFID Signal Strength

- Device-free activity recognition has received a lot of attention lately because it doesn't require participants to wear any gadgets. Instead, sensor devices are placed in the environment, allowing radio signal changes caused by the subject's movements to be recorded and analyzed in order to identify activities [12].
- Recent research [13, 14] suggests utilizing RFID signals for activity recognition without the use of a gadget. These works often include strict guidelines for tag placement, including standards for tag density and tag spacing.

Experimental data and findings

- The trials are carried out using the dataset for healthy older people.
- A sample dataset with consistent activities is the Healthy Older People dataset[1].
- The Healthy Older People dataset (4 RFID reader antennas) experiments are set up with seven groups of diverse male samples as corpus candidates and seven groups of varied female samples as the test set.

Comparison experiments with different models

- In conventional machine learning models, SVM is a representative model. SVM is used as a comparison in this paper and is referred to as SVMI because it only uses independent activity-units.
- A model called incremental learning changes as the data does. This comparison model creates a human activity recognition model while combining incremental learning and voting models.
- In this section, the original LSTM adopts the same research stance as our suggested model. As a result, it also works as a comparison experiment in this case.

Table 10 Accuracy of different comparison models in healthy older people dataset

	SVMI	ANNI	IL+Vote	LSTM	Transformer	BERT	DCO	DCW
1	0.762	1	0.93	1	0.931	0.966	0.897	1
2	0.0674	1	0.0278	0	0	0	0	0
3	1	0	0.0137	1	1	1	0.108	0.892
4	0.167	0.833	0.143	0.857	0.857	0	0	1
Average	0.499	0.708	0.279	0.667	0.697	0.491	0.251	0.723

(Table 10 data collected from [1])

- The most popular models in NLP right now are the Transformer and BERT. These two models are utilized as comparison models because the strategy in this work is an NLP exploration. The paper exclusively uses the original TF-IDF for comparison purposes. In every dataset, it performs poorly.

Abbreviations of Experiments

- **SVMI** is the support vector machine with independent activity-units

- A neural network having independent activity units is called an **ANNI**.
- **IL+Vote** is the incremental learning along with manual parameter voting
- **LSTM** is the long short-term memory
- Transformers' Bidirectional Encoder Representation is called BERT
- The wavelet coefficient activity combination with the TF-IDF is **DCO**
- **DCW** is the wavelet coefficient features; MiTAR with activity combinations

Limitations

The sample size and sampling frequency for this paper must meet particular criteria because it is based on the NLP model. On incredibly small samples, the present model in our work is less precise. The first explanation is that various activities have varied timeframes. Short samples may not have the same density of activity units and activity combinations as the existing corpus. All of those have an impact on recognition accuracy. Another reason is that a smaller sample size might not include enough information. It is simple to confuse people. The solution to this issue can ultimately depend on feature engineering.

Conclusion and Future Work

The focus of the research is on activities from a microscopic perspective, and the SVM and NN models utilized in this paper—which each have independent activity-units—fail to capture the significance and consistency among the data units. Although IL+Vote is a good model, it lacks universality due to its strict requirements for manually-set parameters. The Long Short-Term Memory is a good model, but the original Long Short-Term Memory functions poorly; a better model might perform even better. This research's upcoming topic is this. The experimental findings indicate that both Transformer and BERT suffer from a lack of time efficiency, and that the way described in this paper's method does not produce better recognition results. The Transformer and BERT both have complicated end-to-end models as their foundations, which explains why. Although they offer significant benefits in the extensive processing of natural language processing, the corpus made up of activities is not as rich as natural language for the topic of this research. This study investigates how visually challenged individuals can use natural language data analysis to recognize human behavior.

It focuses on the activity from a tiny perspective.

In the HAR field, other NLP algorithms worth studying include the word bag model, several upgraded neural networks (RNN, CNN, etc.), the seq2seq model, the fastText model, etc. The outcome of NLP-based human activity recognition will be significantly improved by the use of more scientific algorithms.

We will investigate this subject in more detail in the future. The primary research focus is on improving the model using a variety of techniques to meet the goal of adapting to activity recognition using a variety of distinct human being samples.

REFERENCES

1. Men, H., Wang, B. & Wu, G. MiTAR: a study on human activity recognition based on NLP with microscopic perspective. *Front. Comput. Sci.* 15, 155330 (2021).
2. Wu, Haitao et al. "Human activity recognition based on the combined SVM&HMM." *2014 IEEE International Conference on Information and Automation (ICIA)* (2014): 219-224.
3. Anguita, D., Ghio, A., Oneto, L., Parra, X., Reyes-Ortiz, J.L. (2012). Human Activity Recognition on Smartphones Using a Multiclass Hardware-Friendly Support Vector Machine. In: Bravo, J., Hervás, R., Rodríguez, M. (eds) Ambient Assisted Living and Home Care. IWAAL 2012, 216–223.
4. Krishnan R, Subedar M, Tickoo O. Bar: Bayesian activity recognition using variational inference. In: Proceedings of the 3rd Workshop on Bayesian Deep Learning. 2018, 1–8
5. Dave V S, Zhang B, Chen P, Hasan M A. Neural-brane: neural bayesian personalized ranking for attributed network embedding. Data Science and Engineering, 2019, 4(2): 119–131
6. Yuan M, Chen E, Lei G. Posture selection based on two-layer AP with application to human action recognition using HMM. In: Proceedings of IEEE International Symposium on Multimedia. 2017, 359–364
7. Xie X. Human action recognition in the range of Wi-Fi with CNN and ELM. Master Thesis, Beijing, University of Posts and Telecommunication,2018
8. Khanna R, Awad M. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. Berkeley California: Apress, 2015
9. Bharti P, De D, Chellappan S, Das S K. Human: complex activity recognition with multi-modal multi-positional body sensing. IEEE Transactions on Mobile Computing, 2018, 18(4): 857–870
10. Stolke A, Omohundro S. Hidden markrov model induction by bayesian model merging. In: Proceedings of the 5th International Conference on Neural Information Processing Systems. 1992, 11–18
11. Khair E L, Ibrahim A. TF*IDF. Boston: Springer US, 2009
12. L. Yao *et al*., "Compressive Representation for Device-Free Activity Recognition with Passive RFID Signal Strength," in *IEEE Transactions on Mobile Computing*, vol. 17, no. 2, pp. 293-306, 1 Feb. 2018
13. D. Zhang, J. Zhou, M. Guo, J. Cao, and T. Li, "Tasa: Tag-free activity sensing using rfid tag arrays," IEEE Trans. on Parallel and Distributed Systems, vol. 22, no. 4, pp. 558–570, 2011.
14. J. Han, C. Qian, D. Ma, X. Wang, J. Zhao, P. Zhang, W. Xi, and Z. Jiang, "Twins: device-free object tracking using passive tags," in Proceedings of IEEE International Conference on Computer Communications, 2014.