A Robust Sign Language Recognition System with Sparsely Labeled Instances Using Wi-Fi Signals

Jiacheng Shang, Jie Wu

Center for Networked Computing

Dept. of Computer and Info. Sciences

Temple University

Motivation

- Wi-Fi signals are available almost everywhere.
- Wi-Fi signals can monitor surrounding activities using Channel State Information (CSI).

Motivation

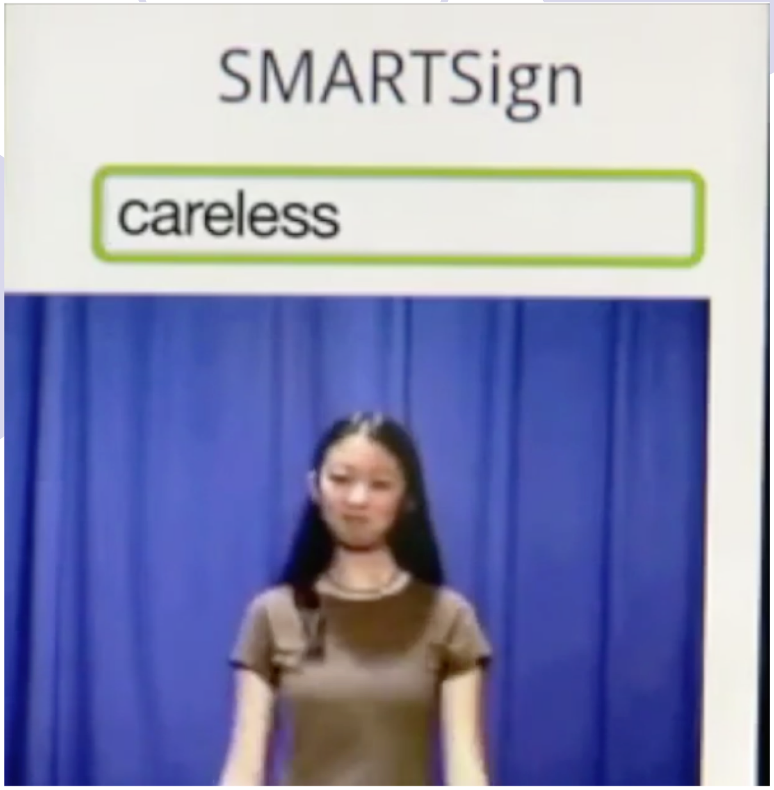
- Sign language (SL) mainly uses manual communication to convey meaning.

Motivation

- Automatic SL Recognition is still in its infancy.
- Currently, all commercial translation services are human-based, and therefore, expensive.

Motivation

- Automatic SL Recognition is still in its infancy.
- Currently, all commercial translation services are human-based, and therefore, expensive.

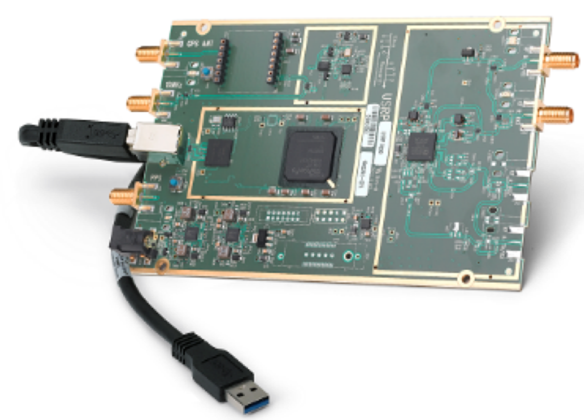


Motivation

- Automatic SL Recognition is still in its infancy.
- Currently, all commercial translation services are human-based, and therefore, expensive.
- American Language Services offers interpreters starting at $125 per hour and a two-hour minimum is required

Problem Statement

- Sign language recognition using Wi-Fi signals
  - Uses commercial Wi-Fi devices (routers and laptops) to recognize sign language.
- Strengths
  - Can work in the dark
  - Avoids breaching user privacy
  - No need to wear sensors
  - Low cost

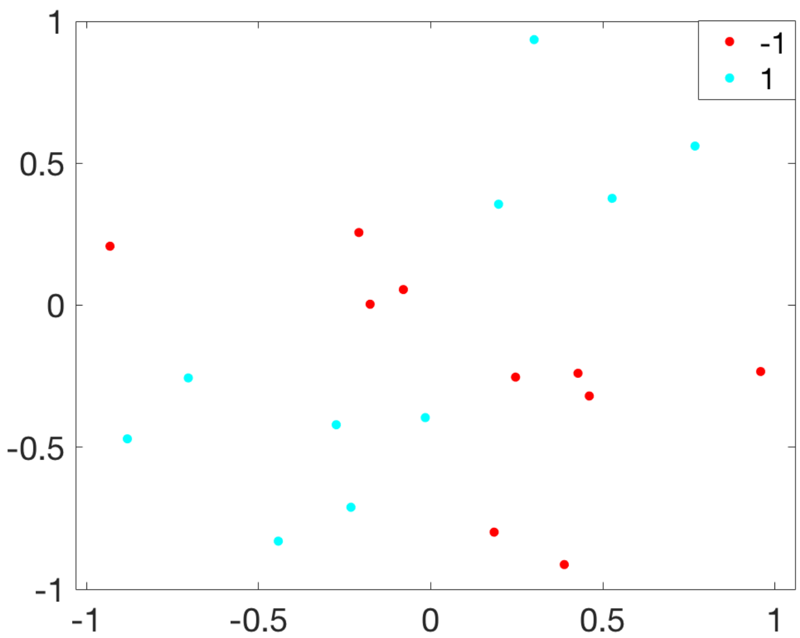


Limitations of Existing Systems

- Limitations of existing systems: models are trained based on a large dataset
  - Large training datasets are usually hard and expensive to get in practice.
  - Many works have the potential requirement that label distributions in the training dataset and the testing dataset should be the same.
- Our approach: reduce the size of the training dataset by leveraging the knowledge in the unlabeled dataset and others' training datasets

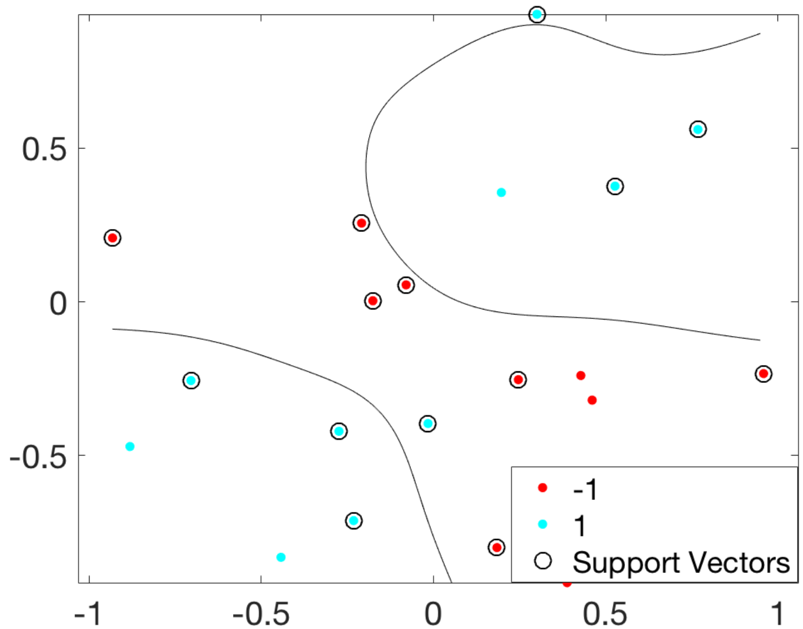
Limitations of Existing Systems

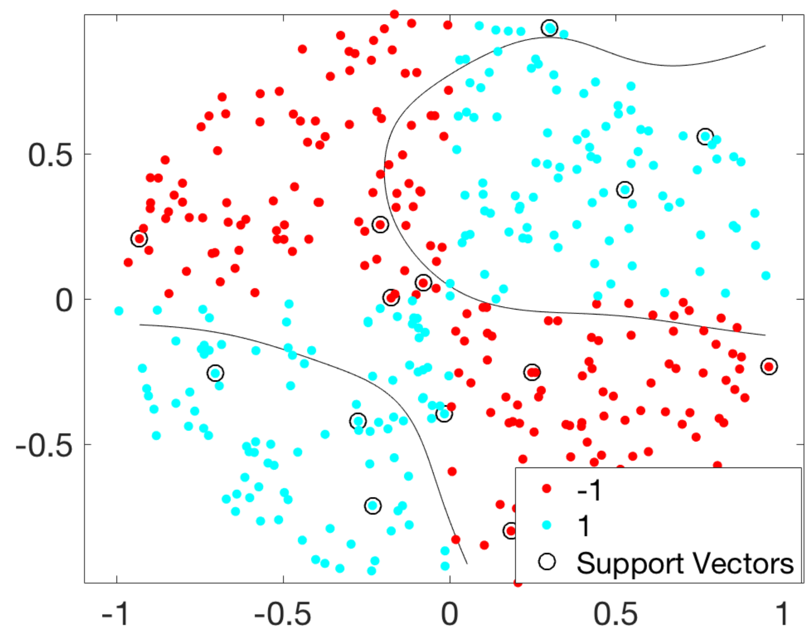
- Why are current models trained using a large dataset?



Limitations of Existing Systems

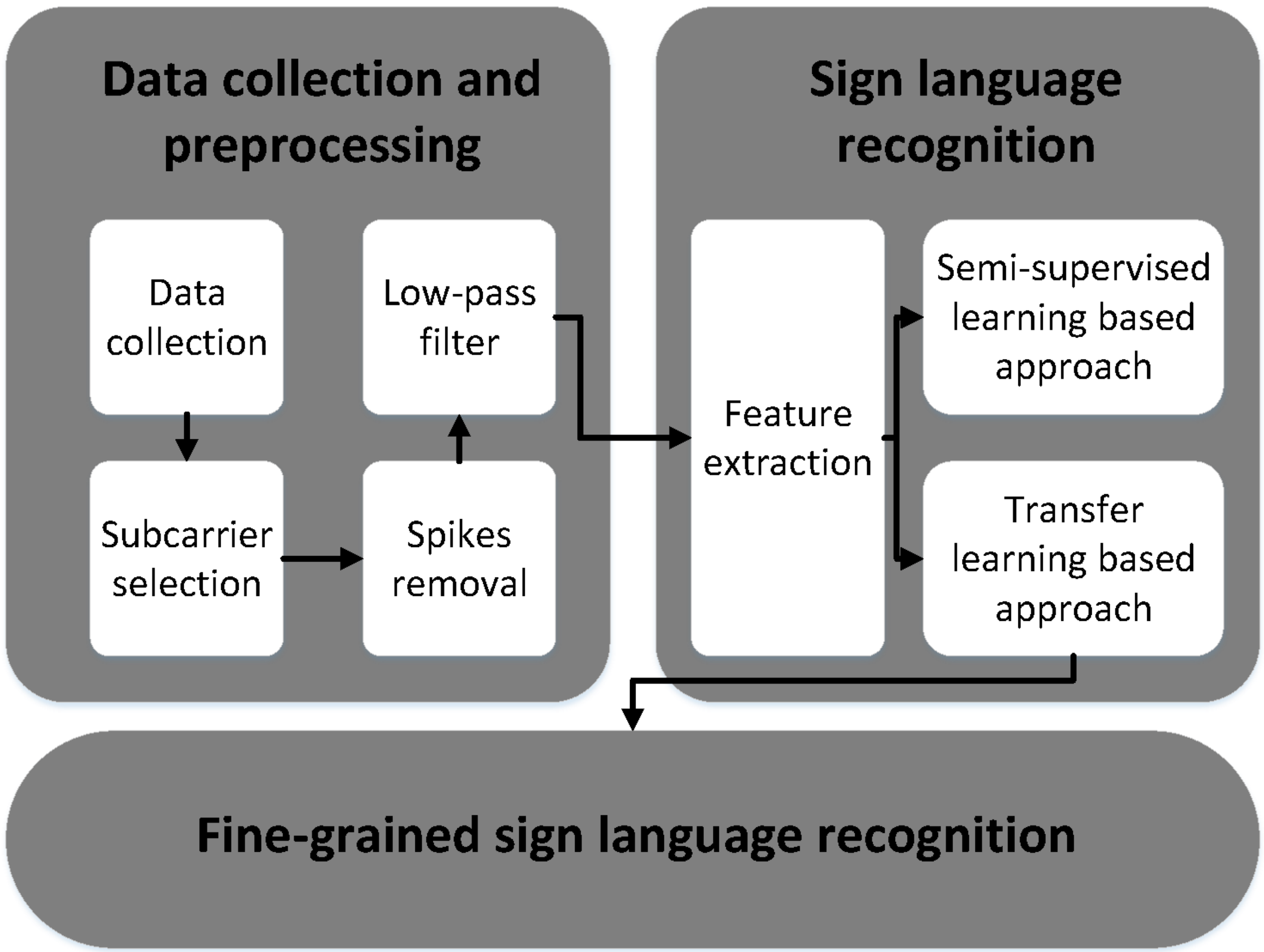
- Why are current models trained using a large dataset?





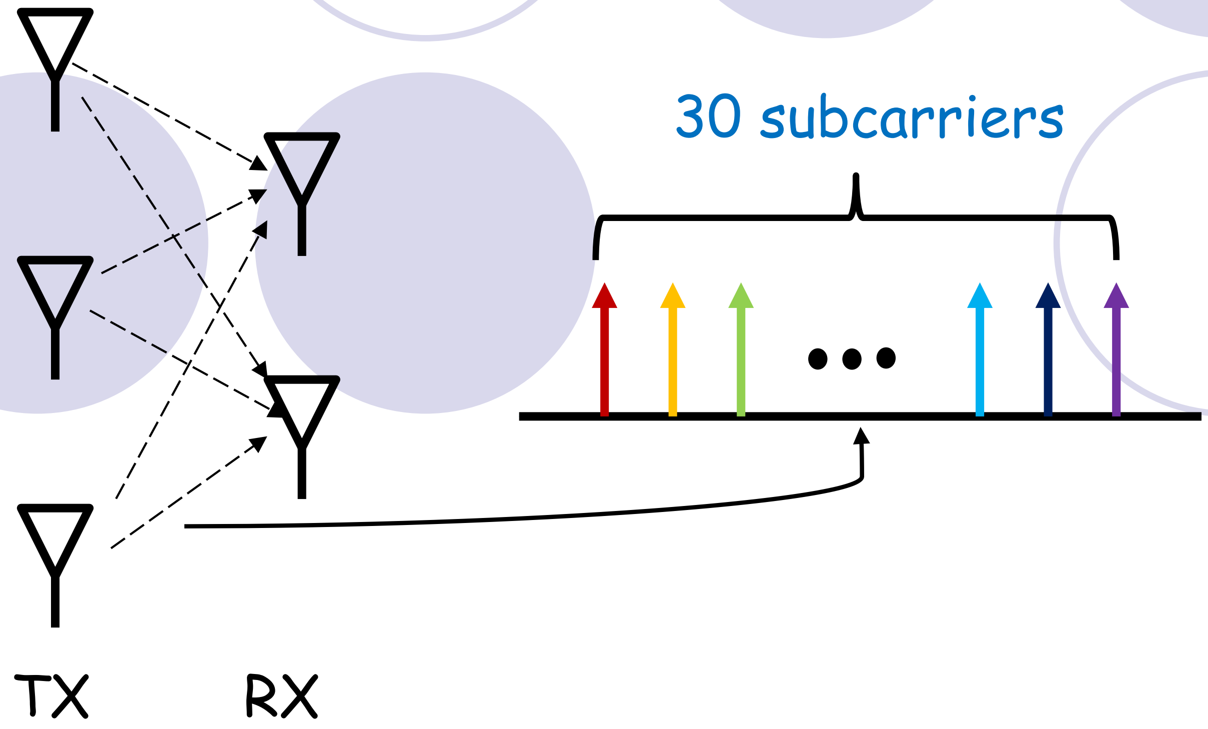
Accuracy: 79%

Sign Language Recognition Pipeline



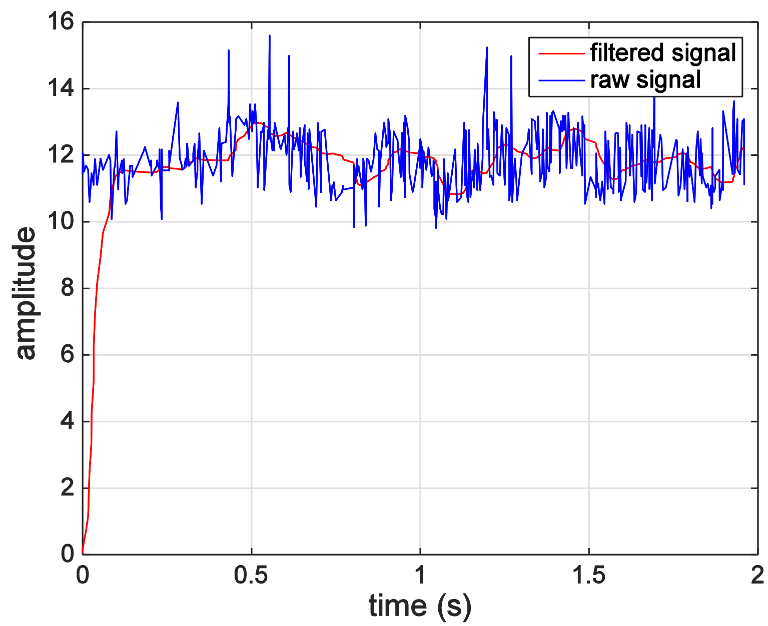
Signal Preprocessing

- Subcarrier Selection
  - Different subcarriers have different sensitivities to different human activities



Signal Preprocessing

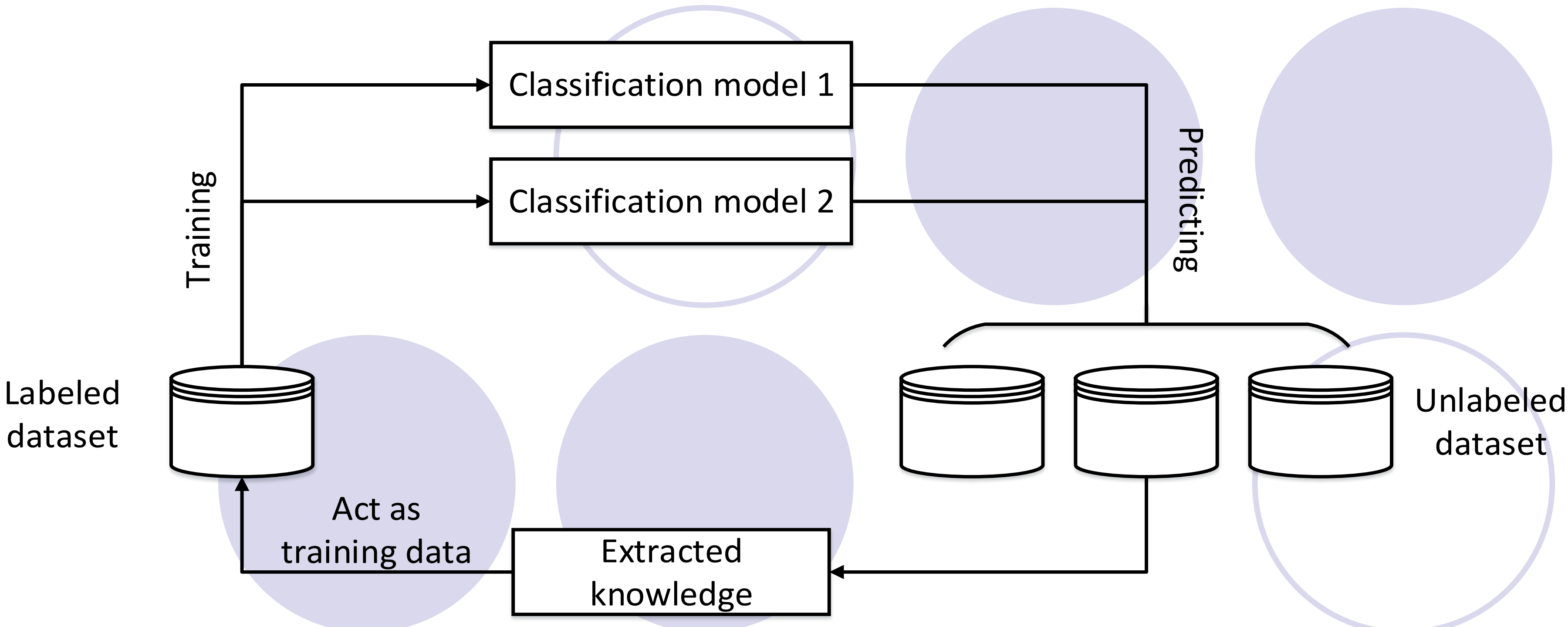
- Noise removal
  - Smoothing: removes outliers
  - Low-pass filter: removes high frequency noise
  - The average amplitude and the average median absolute deviation are chosen as the features.



Leverage knowledge in unlabeled datasets

- Labeled instances are often very time consuming and expensive to obtain.
- The new user may only be able to label some instances while most instances stay unlabeled.
- Knowledge in unlabeled instances can be used to improve the recognition's performance.
- Co-training is an efficient semi-supervised learning paradigm

Leverage knowledge in unlabeled dataset



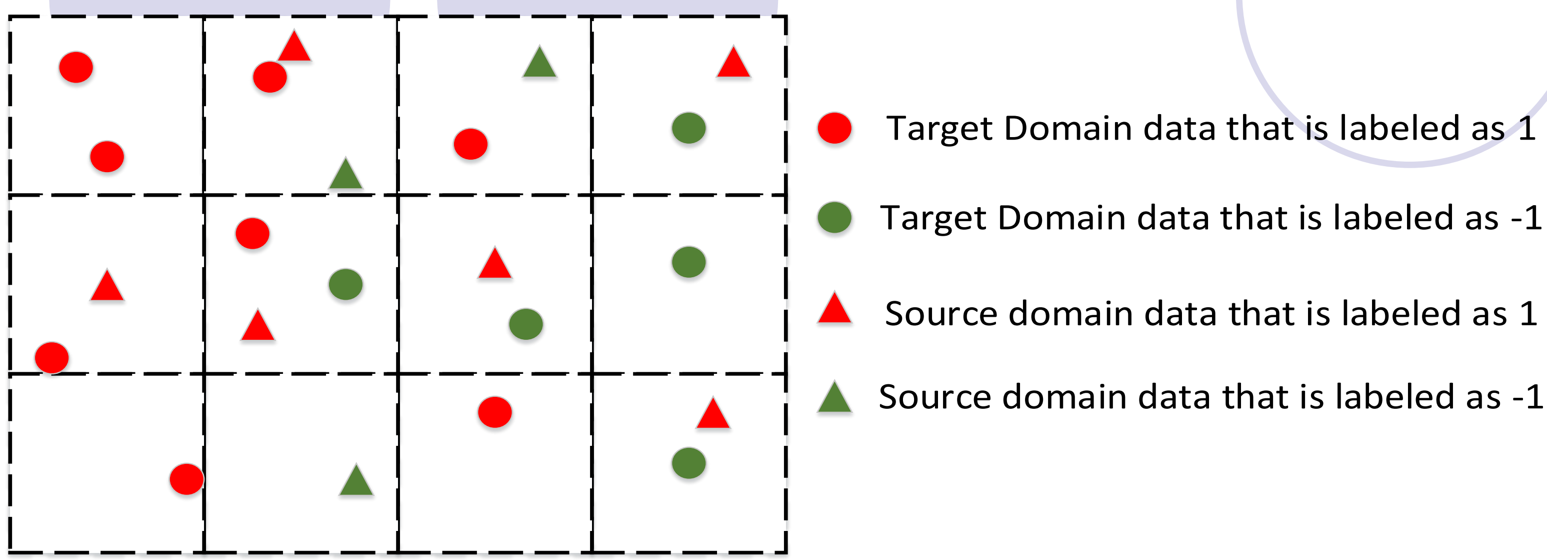
- Extracted knowledge: those unlabeled instances that are predicted as the same label by both (of two) classification models

Reuse others' training datasets

- The ability to recognize and apply knowledge obtained in previous tasks
- Why Reuse?
  - Build every model from scratch?
    - Time consuming and expensive
  - Reuse knowledge extracted from existing tasks and datasets
    - More practical
- How can we decide what data should be transferred to the new user?

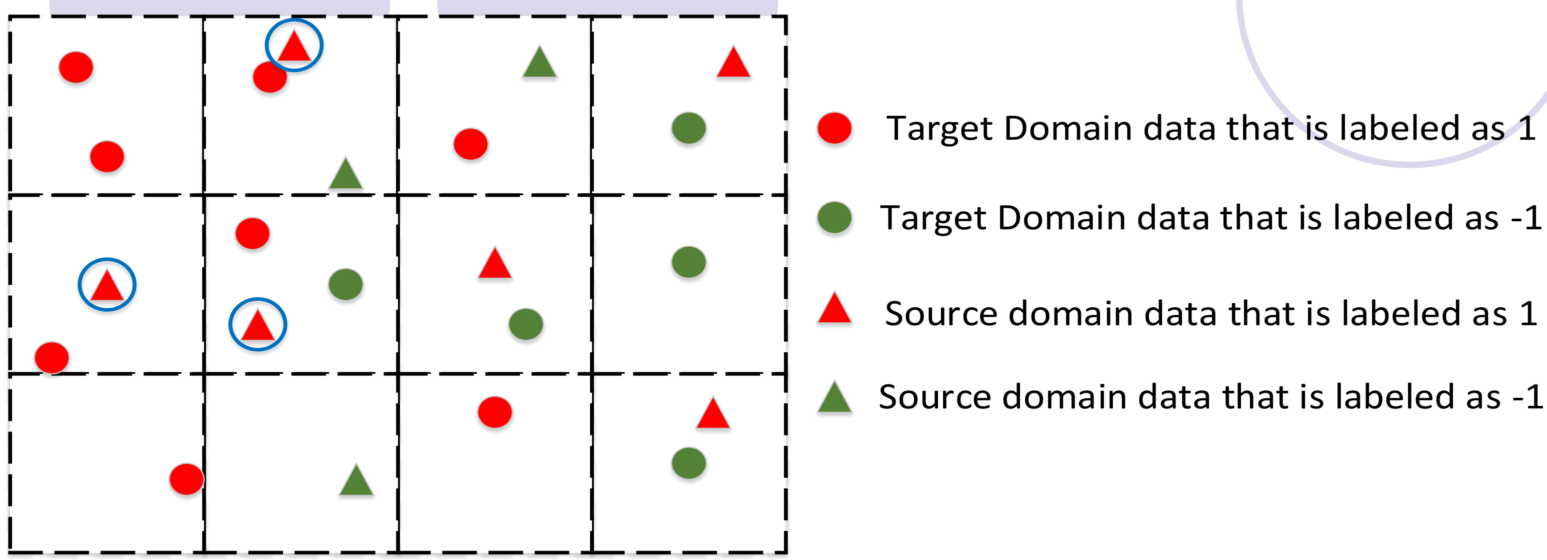
Reuse others' training dataset

- Transfer algorithm: find those useful instances from existing labeled source domain data
  - Features value discretization on each dimension with a grid size of τ.
  - A source domain instance is transferred to target domain iff there is a target domain instance with the same label in the same grid.

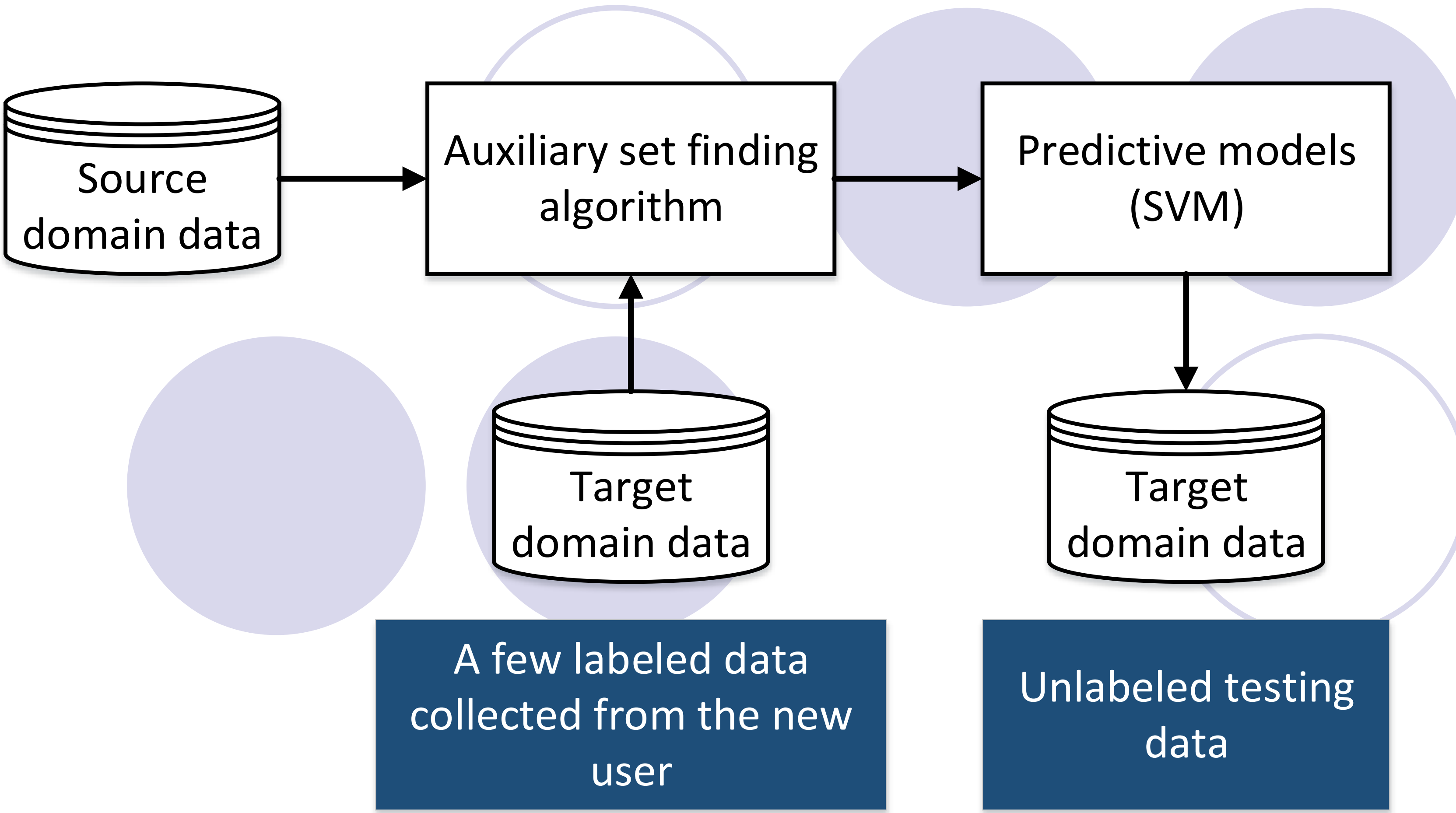


Reuse others' training dataset

- Transfer algorithm: find those useful instances from existing labeled source domain data
  - Features value discretization on each dimension with a grid size of τ.
  - A source domain instance is transferred to target domain iff there is a target domain instance with the same label in the same grid

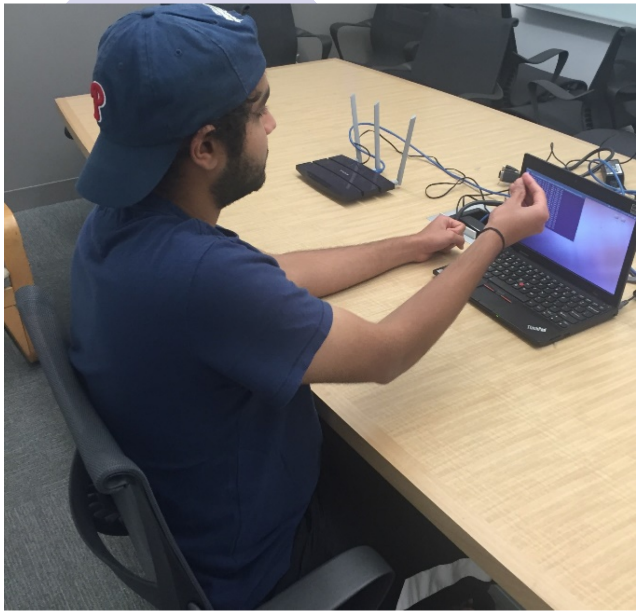


Reuse others' training dataset



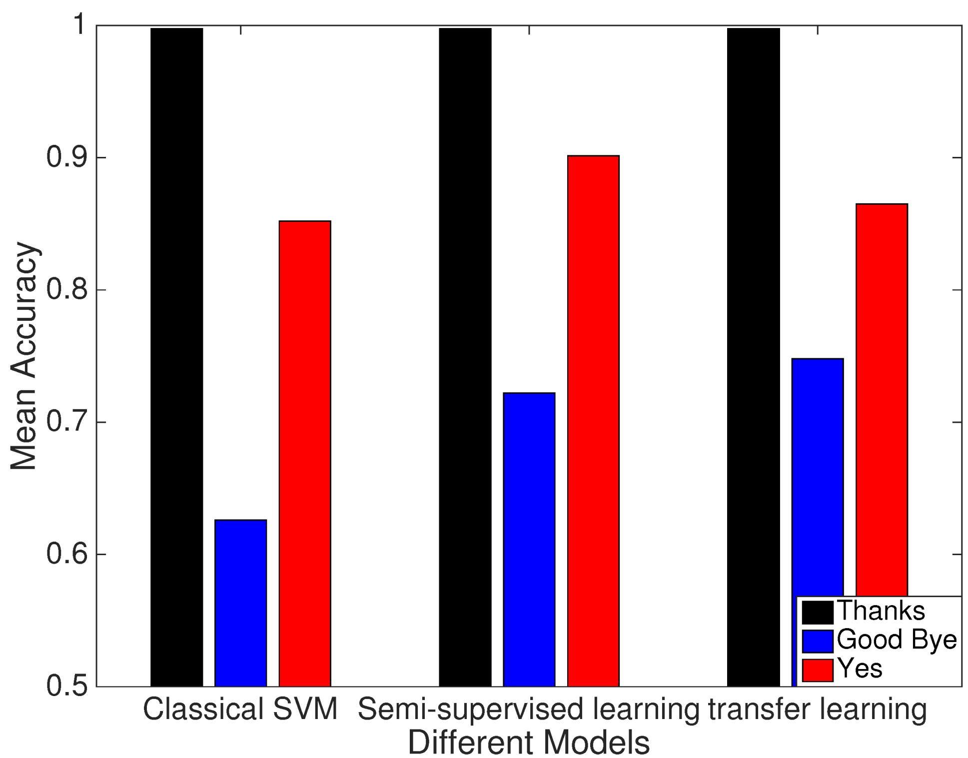
Evaluation

- Commercial hardware with no modifications
  - Transmitter: TP-Link TL-WR1043ND Wi-Fi router
  - Receiver: Lenovo X100e laptop with Intel 5300 NIC
  - Downloading a large file from an FTP server within the same local network area



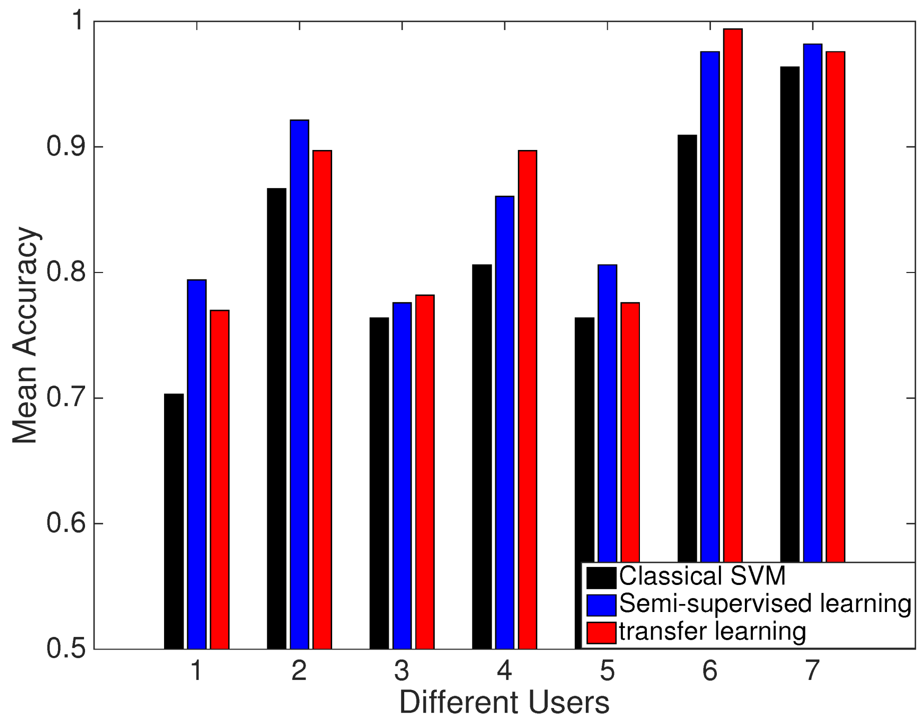
Evaluation Results

- Mean accuracies vs. different solutions
  - Two proposed solutions can achieve better accuracies with sparsely labeled training data.



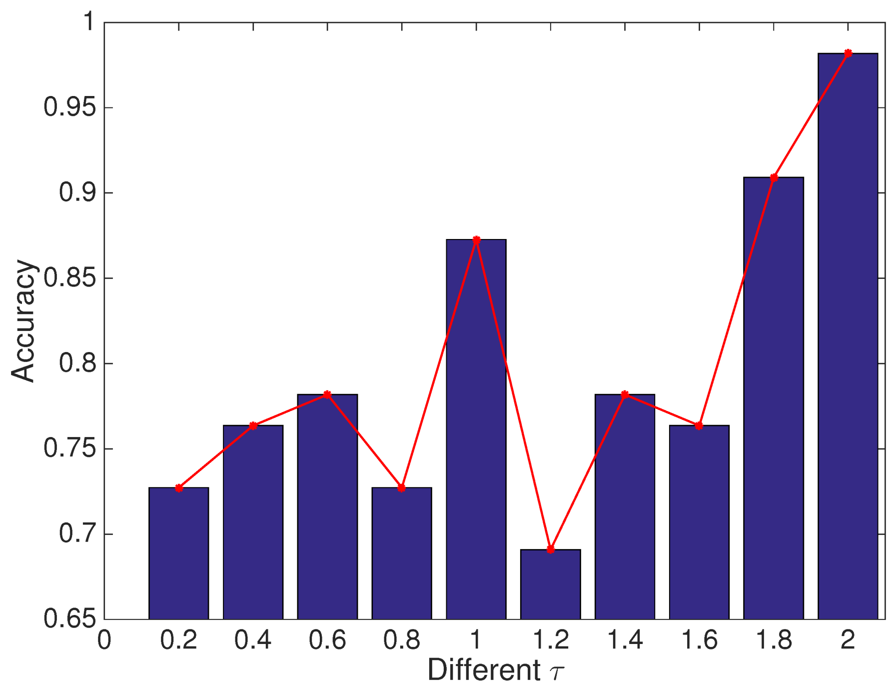
Evaluation Results

- Mean accuracies vs. different users
  - Our approaches can achieve a mean prediction accuracy of about 87% for all participants.



Evaluation Results

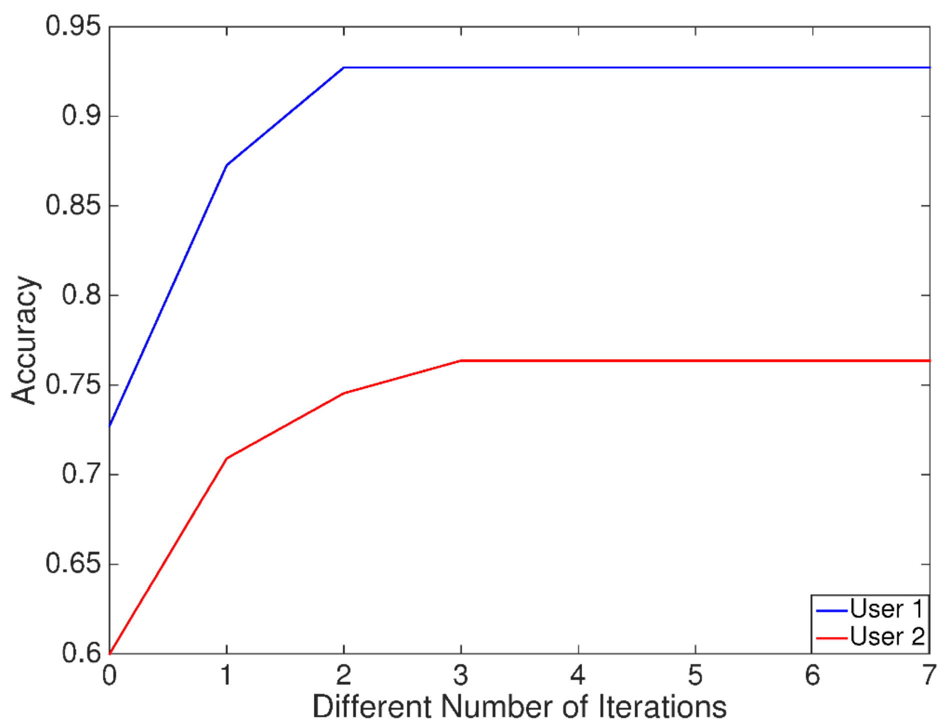
- Accuracies vs. different τ
  - There is no linear relationship between the accuracy and τ.
  - T is determined based on the distribution and density of the data.



MASS 2017

Evaluation Results

- Accuracies vs. different iterations
  - We set the number of iterations to 5 in our system.



Conclusion

- CSI measurements contain fine-grained movement information and can be used to recognize sign language.
- Propose a sign language recognition system that can achieve a good performance with sparsely labeled data.
  - Leveraging the knowledge in an unlabeled dataset.
  - Reusing others' training datasets.
- Experimental results show that our system can achieve a mean prediction accuracy of about 87%.

Thanks!

Q & A