

# PIVE: Per-Iteration Visualization Environment for Real-time Interactions with Dimension Reduction and Clustering

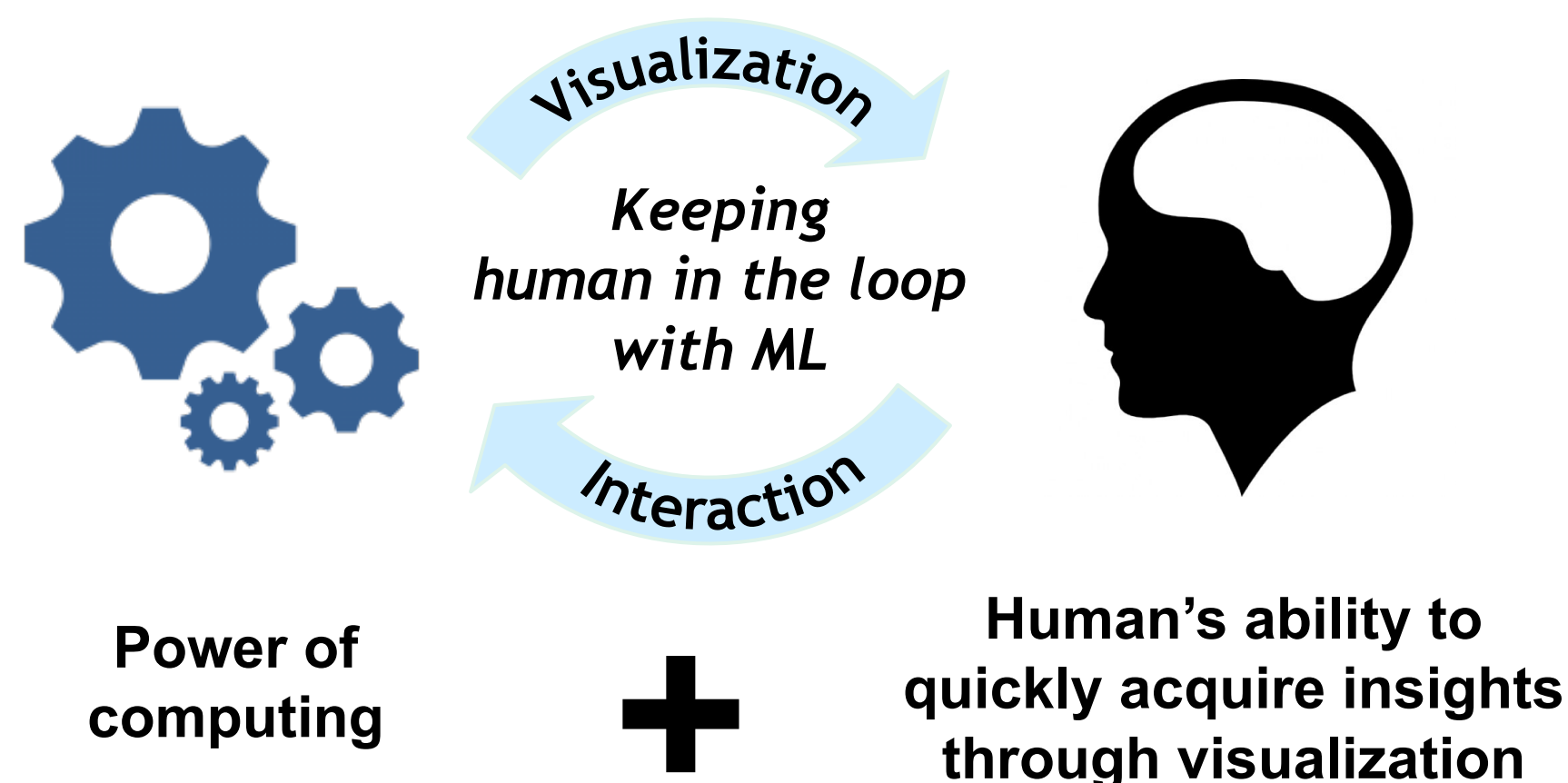
[tiny.cc/aaai17pive](http://tiny.cc/aaai17pive)

Hannah Kim<sup>1</sup>, Jaegul Choo<sup>2</sup>, Changhyun Lee<sup>3</sup>, Hanseung Lee<sup>3</sup>, Chandan K. Reddy<sup>4</sup>, and Haesun Park<sup>1</sup>

<sup>1</sup>Georgia Tech, <sup>2</sup>Korea University, <sup>3</sup>Google Inc., <sup>4</sup>Virginia Tech



## Visual Analytics and Machine Learning

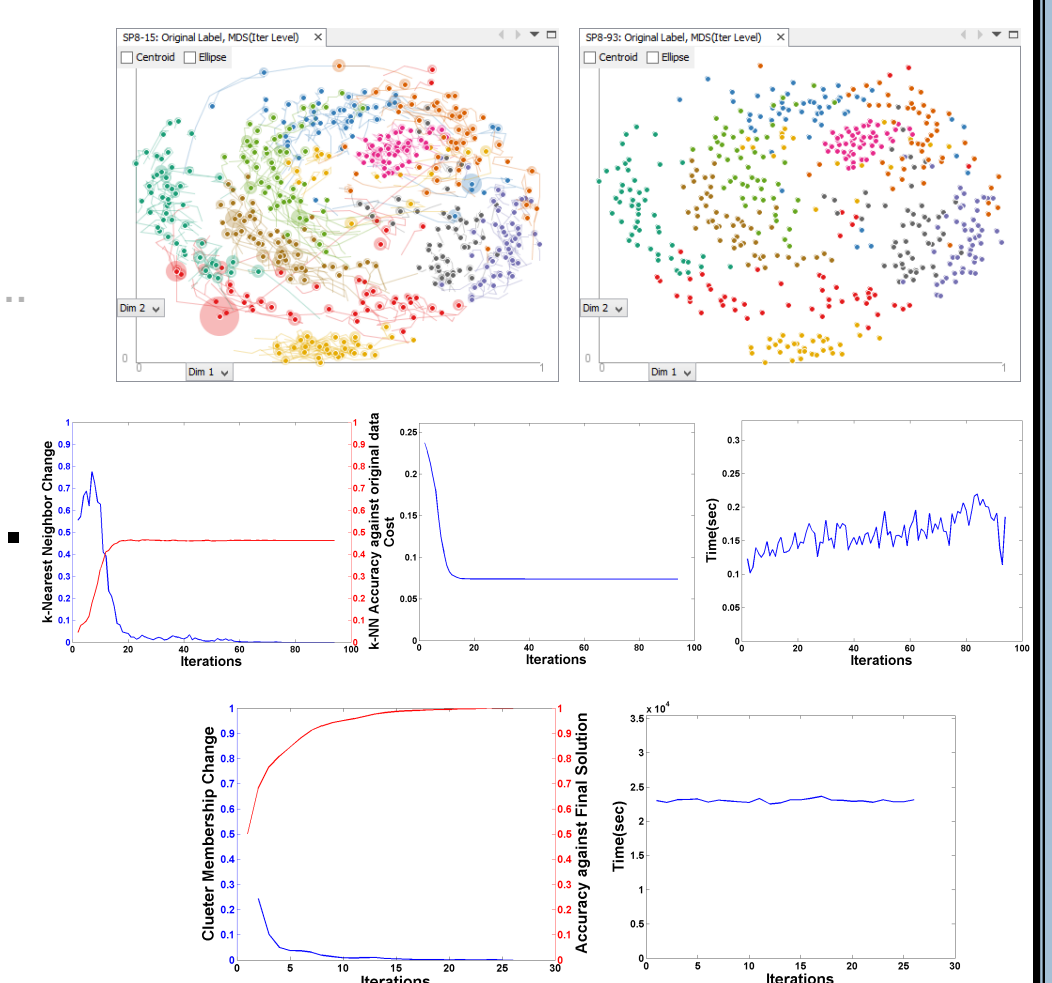


However, as ML methods become more advanced, their high costs hinder real-time interactive visualizations with them.

Even the state-of-the-art in visual analytics adopts only a few standard techniques and does not properly leverage advanced ML.

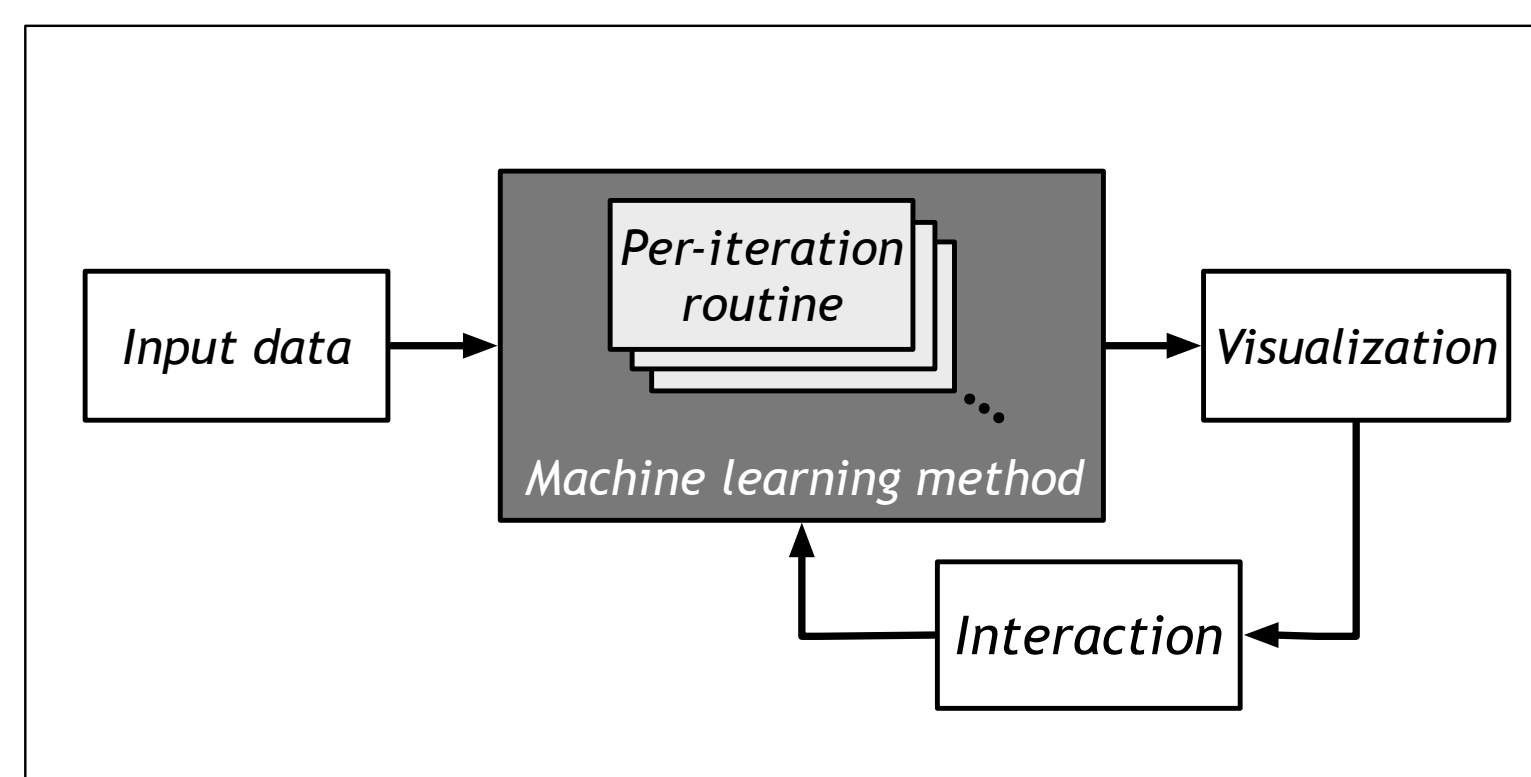
## Motivation

- Humans' perceptual precision**
- Visual perception does not require highly precise outputs, e.g.,  $\pi = 3.14159265359\dots$
- Iterative behavior of machine learning**
- Many computational methods are iterative.
  - A major improvement typically occurs in early iterations.
- **Let's cut through algorithm iterations to save time for interactive visualization.**



## Per-Iteration Visualization Environment (PIVE)

### The standard vs. PIVE workflow

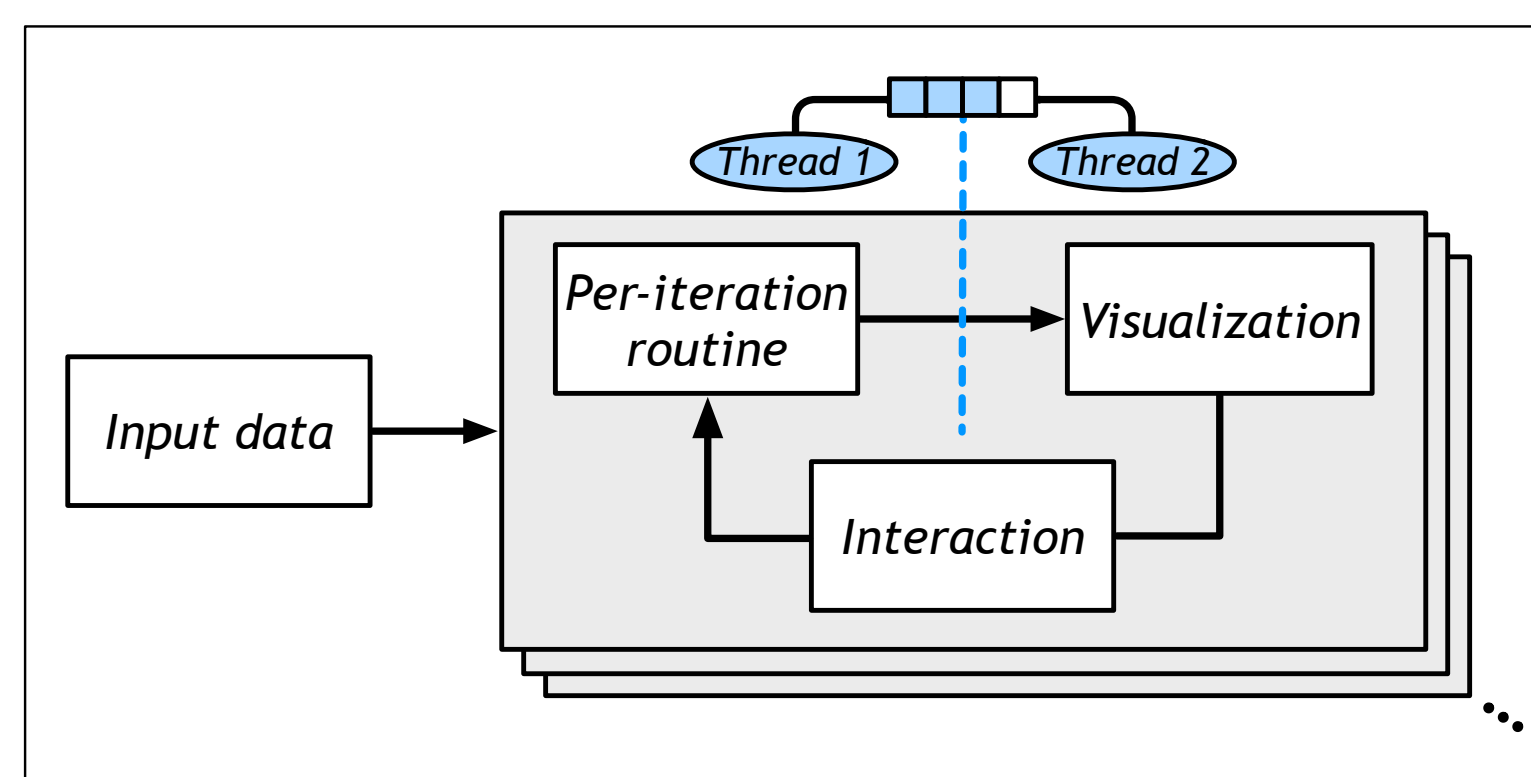


Previously...

- ✗ Treat ML as a black box
- ✗ Wait till algorithms finish

with PIVE

- ✓ Interact with intermediate result
- ✓ No additional cost
- ✓ Applicable to various iterative methods



### Interaction Methodology

: User replaces intermediate output during algorithm iterations.

- 1) Soft replacement: Replaced output works as a new initialization.
- 2) Hard replacement: Replaced output remains same. Viewed as constrained or semi-supervised method.

#### Algorithm 1 Iterative methods

```

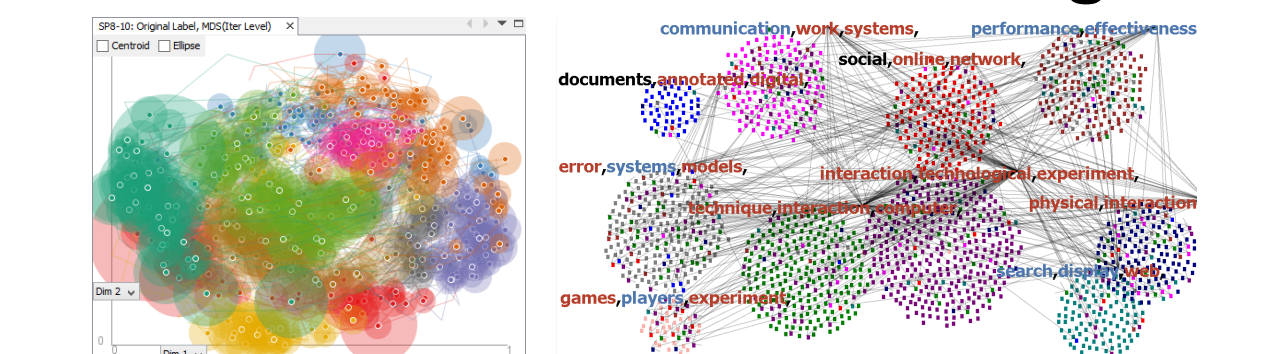
1: Input:  $X = \{x_1, \dots, x_n\}$  and parameter  $\alpha$ 
2: Output:  $Y = \{y_1, \dots, y_n\}$ 
3:  $t \leftarrow 0$ 
4: Initialize  $Y^t = \{y_1^t, \dots, y_n^t\}$ 
5: repeat
6:    $t \leftarrow t + 1$ 
7:   /* Per-iteration routine */
8:   for  $i \leftarrow 1, \dots, n$  do
9:      $y_i^t \leftarrow f(\{X, Y^0, \dots, Y^{t-1}\}, \alpha)$ 
10:   $Y^t \leftarrow \{y_1^t, \dots, y_n^t\}$ 
11: until a stopping criterion is satisfied
12:  $T \leftarrow t$  /* Final iteration index */
13:  $Y \leftarrow Y^T$  /* Final output */
    
```

Replace  $y_i^t \leftarrow \tilde{y}_i$  for this iteration (soft) for all iterations (hard)

### Design Considerations

#### Stability and Convergence

- One needs to know when visualization becomes stable.
- PIVE explicitly visualizes intermediate changes in charts and visual encoding.



#### Computational Overhead

- Constant processing of intermediate output
- We use multithreading (two threads for computation and visualization).

## Supported Interaction

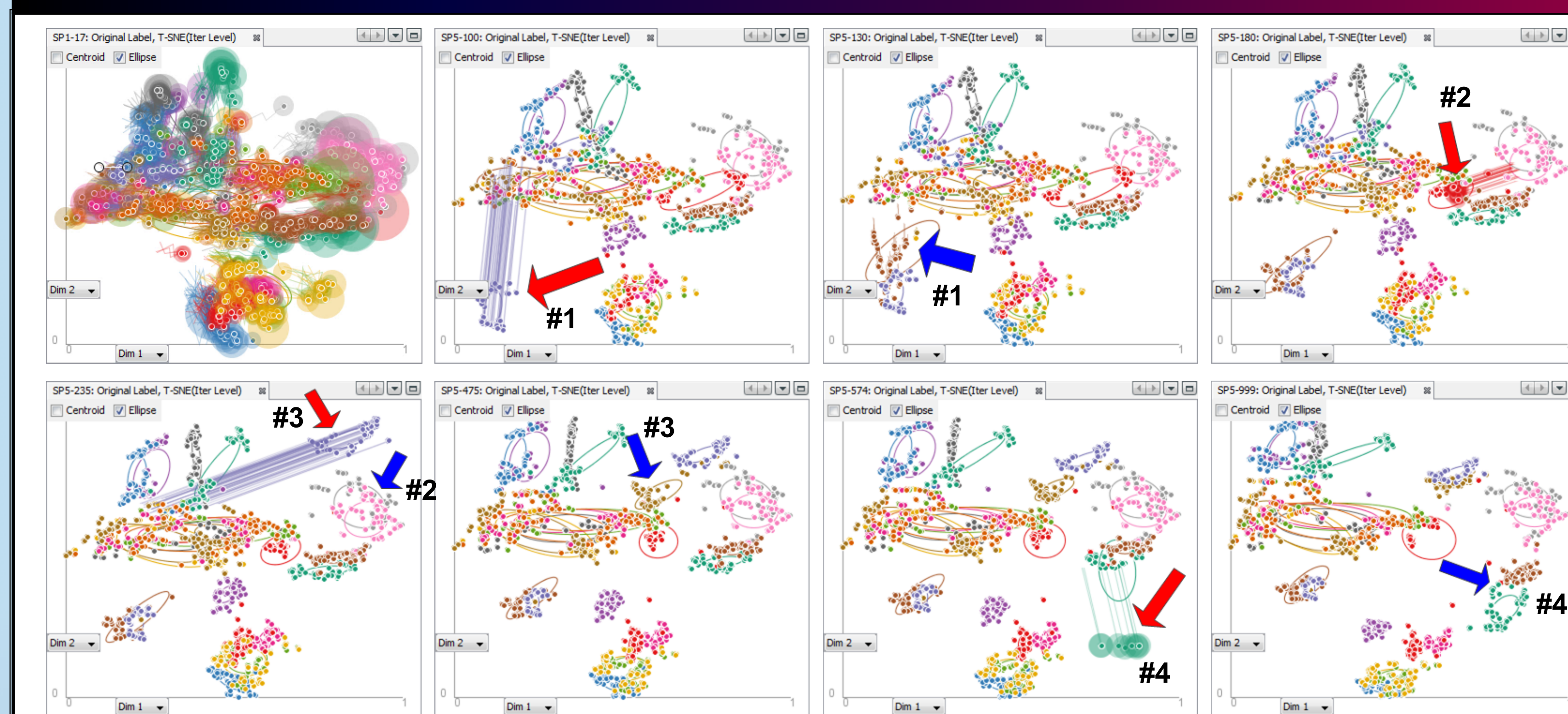
### Dimension Reduction

- Change parameters
- Filter data points
- Move data points

### Clustering

- Change parameters
- Filter data points
- Change membership of data points
- Split/Merge clusters
- Freeze clusters

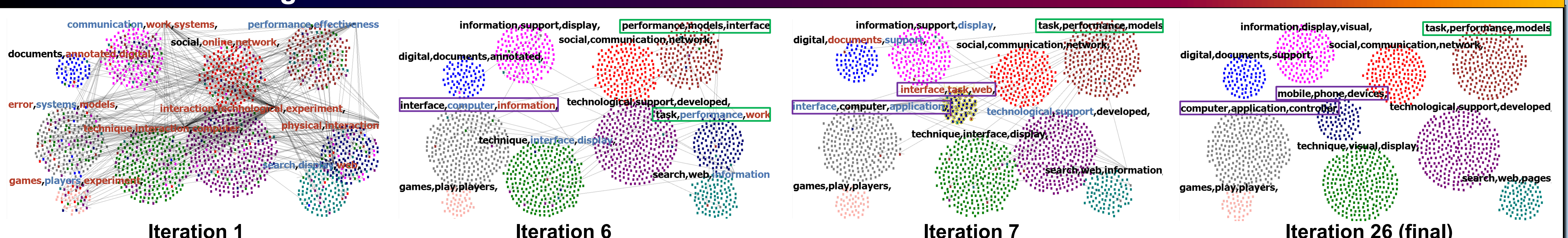
## Interactive Dimension Reduction with PIVE



Hard replacement interactions with t-SNE. Red (and blue) arrows point to interactions (and results). 1,558 spoken letter data represented in 618 dimensions have been used.

- 1) Remove 'c' from overlapping clusters → Similar 'z' follows
- 2) Remove part of 'w' from overlapping clusters → 'm' and 'n' stays
- 3) Remove 'q' from overlap → 'u' follows, '-ee' stays
- 4) Separate 'l' and 'o' → clearer

## Interactive Clustering with PIVE



Hard replacement interactions with k-means. At the sixth iteration, we perform cluster merging (green rectangles) and splitting (a purple rectangle) interactions, and the final result due to these interactions is converged in 26th iterations. 1,671 CHI conference papers published between 1999 and 2010 have been used.