Integrating Specialized Classifiers Based on Continuous Time Markov Chain

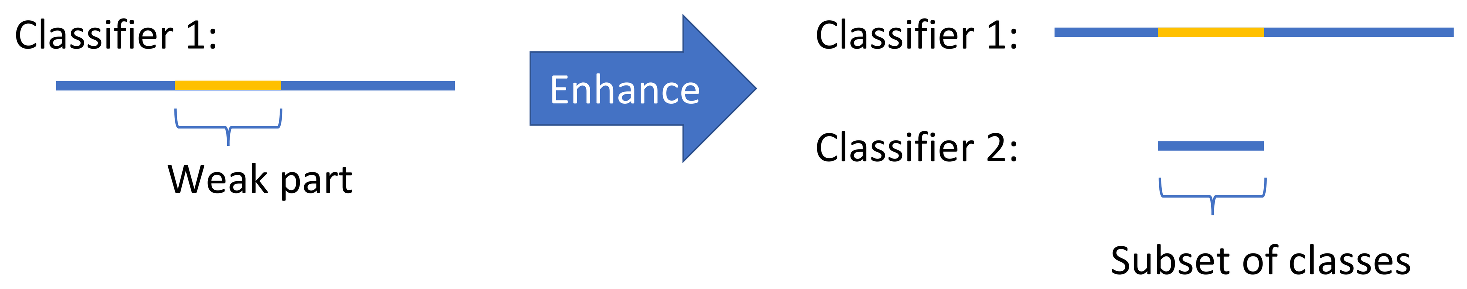
Zhizhong Li, Dahua Lin; *Multimedia Laboratory, The Chinese University of Hong Kong*

香港中文大學 The Chinese University of Hong Kong

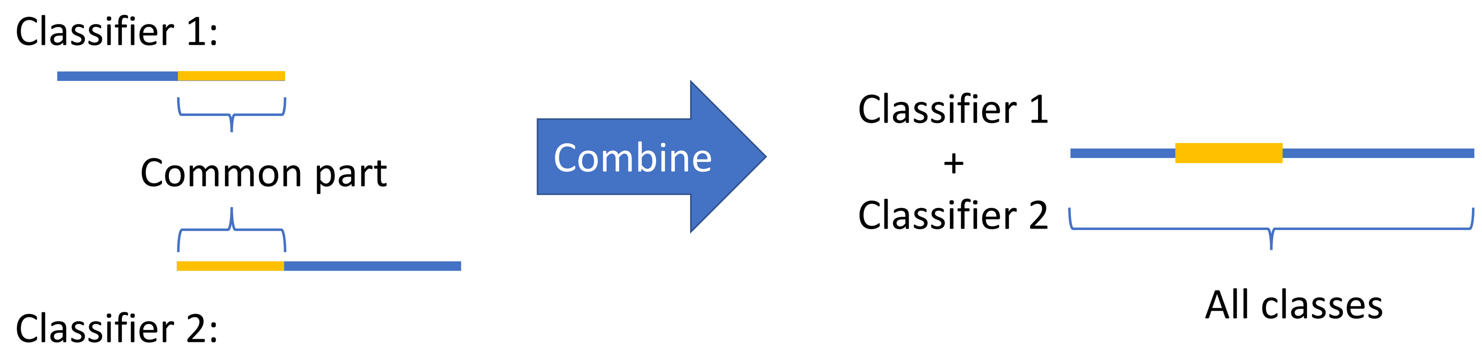
IJCAI-17 MELBOURNE

Ensemble of Specialized Classifiers

- We consider ensemble in the presence of Specialized Classifiers
- **Specialized classifiers** are trained using a subset of classes
- Scenario 1: Enhance the weak part



- Scenario 2: Integrate classifiers from different domain

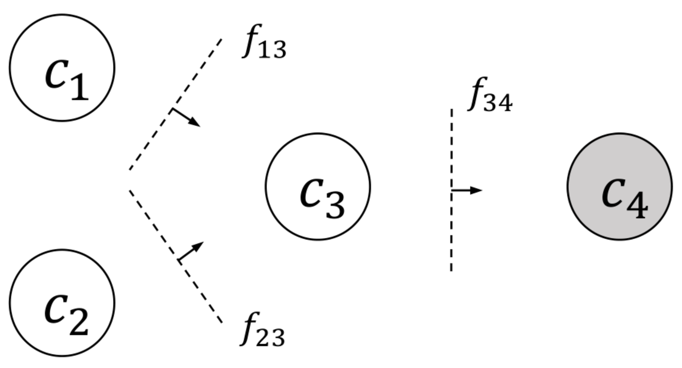


- **Challenges**

From local to global

Unbalanced coverage

- **Motivation**: tracing the preferences



4 classes in 2-dim plane. 3 specialized classifiers are drawn with decision boundaries
If we trace along the arrows, we will arrive the correct class.

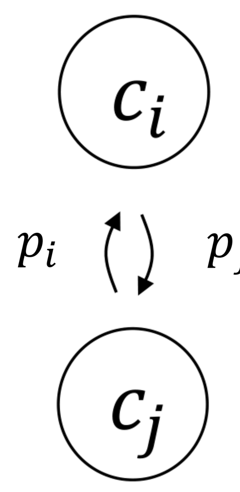
- **Idea**: tracing -> random walk
  If the configuration of specialized classifiers are complex, the simple tracing algorithm faces problems like multiple solutions, or trapped in loops. Random walk is a better replacement.

Tool: Continuous Time Markov Chain

- Homogeneous CTMC is a stochastic process characterized by the transition rates between states
- If the chain is ergodic, there exists an equilibrium distribution

- **Example**
  Two states
  Transition rate matrix

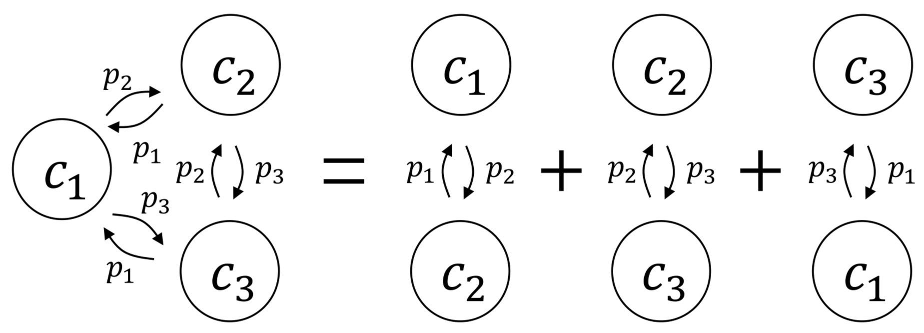
$$\mathbf{Q} = \begin{bmatrix} -p_j & p_j \\ p_i & -p_i \end{bmatrix}$$



- Stationary distribution $\pi \propto (p_i, p_j)$
- A model of preference

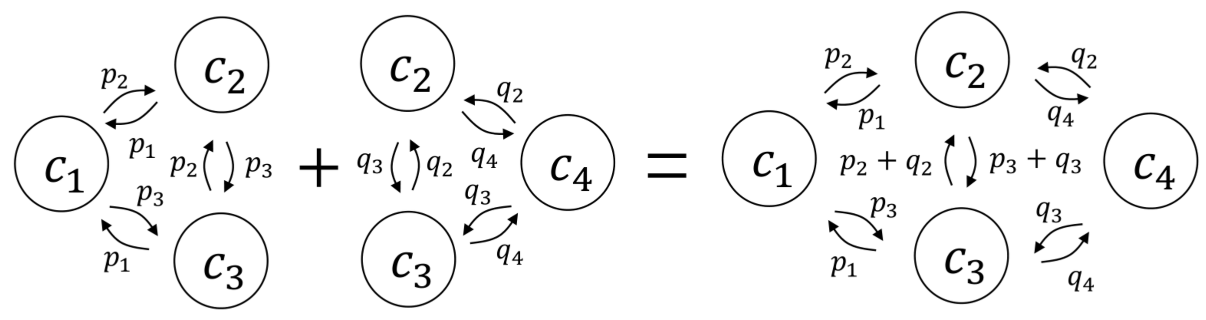
CTMC Ensemble

- Given a set of classifiers, assume they output probabilities
- Construct a continuous-time Markov chain as follows
- **Classes as states, predictions as transition rates**
- **Step 1**: Construct a chain for each of individual classifiers



Three classes with prediction (p_1, p_2, p_3),
Stationary distribution is also (p_1, p_2, p_3).

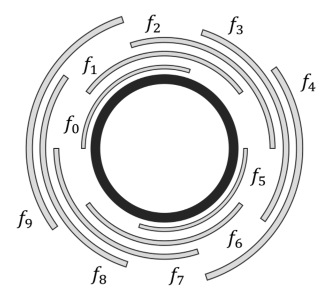
- **Step 2**: Combine the by superposition



- **Step 3**: Output the equilibrium distribution as prediction

Experiments

- On ImageNet dataset
  1000 classes, 1.2M images
  Train deep neural networks as classifiers



- From local to global

Classifier Set	Top-5 Error (%) vote	mean	m-mean	prod	m-prod	CTMC
$\{f_0, f_2, f_4, f_6, f_8\}$	15.962	12.332	11.734	13.376	11.144	**7.952**
$\{f_1, f_3, f_5, f_7, f_9\}$	16.052	12.742	12.078	12.600	10.686	**7.718**
$\{f_0, f_2, f_4, f_6, f_8\} \cup \{f_1\}$	19.706	12.148	11.508	12.964	10.536	**7.696**
$\{f_0, f_2, f_4, f_6, f_8\} \cup \{f_1, f_3\}$	23.898	11.806	11.046	12.924	10.050	**7.528**
$\{f_0, f_2, f_4, f_6, f_8\} \cup \{f_1, f_3, f_5\}$	22.020	11.096	10.620	12.326	9.634	**7.206**
$\{f_0, f_2, f_4, f_6, f_8\} \cup \{f_1, f_3, f_5, f_7\}$	14.856	10.334	10.136	9.954	9.206	**6.886**
$\{f_0, f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9\}$	14.172	10.014	10.014	8.548	8.548	**6.676**

- Unbalanced coverage

Classifier Set	Top-5 Error (%) vote	mean	m-mean	prod	m-prod	CTMC
$\{h_0\}$	21.998	6.146	6.146	6.146	6.146	6.146
$\{h_0\} \cup R$	16.540	8.356	8.356	6.404	6.404	**5.972**
$\{h_0\} \cup S$	22.600	15.292	15.292	7.670	7.670	**6.108**
$\{h_0\} \cup U$	25.912	16.532	15.196	9.948	8.794	**5.802**

h_0: general classifier. R, S, U: three sets of specialized classifiers.
Classes are covered by 1-4 times unevenly.

Properties

- **Proposition 1**. If the coverage is connected, and the prediction probabilities are positive, then the solution **exists** and is **unique**
- **Proposition 2**. The output π satisfies

$$\pi(i) = \frac{1}{|\mathcal{F}_i|} \sum_{k \in \mathcal{F}_i} \omega_k \mathbf{p}_k(i), \quad with \ \omega_k = \sum_{j \in I_k} \pi(j).$$

where $\mathcal{F}_i$ is the set of all classifiers that covers class c_i, and $\mathbf{p}_k(i)$ is the prediction of k-th classifier on class c_i.

Codes available at https://github.com/innerlee/CTMCEnsemble.jl