Multi-core Structural SVM Training

Kai-Wei Chang, Vivek Srikumar and Dan Roth

Motivation

Many applications require structured decisions. Global decisions where local decisions play a role but there are mutual dependencies on their outcome. It is essential to make coherent decisions in a way that takes into account the interdependencies among different factors.

Part-of-Speech tagging (sequential labeling)

Input: a sequence of words \( x_1, x_2, \ldots, x_n \).

Output: POS tags \( y_1, y_2, \ldots, y_n \), \( y_i \in \{NN, VBZ, \ldots\} \)

“A cat chases a mouse” => “DT NN VBZ DT NN”.

Assignment to \( y_i \) can depend on both \( x_i \) and \( y_{i-1} \).

Feature vector \( \phi(x, y) \) defined on both input and output variables: e.g., “x; Cat; “, “y1: VBZ”.

Structured Prediction Model

Structured prediction: predicting a structured output variable \( y \) based on the input variable \( x \).

\( y = (y_1, y_2, \ldots, y_n) \) variables form a structure: sequences, clusters, trees, or arbitrary graphs.

Various approaches have been proposed to learn structured prediction models [Machnies et. al. 09, Chang and Yih 13, Lacoste-Julien et. al. 13] but they are single-threaded.

DEMI-DCD: a multi-threaded algorithm for training structural SVM.

Advantages:

- Requires little synchronization between threads.
- Fully utilizes the power of multiple cores.
- Makes multiple updates on the structures discovered by the loss-augmented inference.
- Fully utilizing the available information.

Inference

\[ \arg \max_{y \in Y} \phi(x, y) \]

Weight parameters.

Set of allowed structures often specified by constraints.

Features on input-output.

Efficient inference algorithms have been proposed for some specific structures.

Integer linear programming (ILP) solver can deal with general structures.

Abstract

Many problems can be framed as structured prediction problems. Structural support vector machines (SVM) is a popular approach for training structured predictors. In structured SVM learning, inference is done by alternating between an inference (prediction) phase and a model update phase. The inference phase selects candidate structures for all training examples and then the model is updated based on these structures. This paper develops an efficient multi-core implementation for structural SVM. We extend the dual coordinate descent approach by decoupling the model update and inference phases into different threads. We prove that our algorithm not only converges but also fully utilizes all available processors to speed up learning.

Motivation:

Inference

Output

Input

Efficient

Structured

Motivation:

Many applications require structured decisions. Global decisions where local decisions play a role but there are mutual dependencies on their outcome. It is essential to make coherent decisions in a way that takes into account the interdependencies among different factors.

Part-of-Speech tagging (sequential labeling)

Input: a sequence of words \( x_1, x_2, \ldots, x_n \).

Output: POS tags \( y_1, y_2, \ldots, y_n \), \( y_i \in \{NN, VBZ, \ldots\} \)

“A cat chases a mouse” => “DT NN VBZ DT NN”.

Assignment to \( y_i \) can depend on both \( x_i \) and \( y_{i-1} \).

Feature vector \( \phi(x, y) \) defined on both input and output variables: e.g., “x; Cat; “, “y1: VBZ”.

Structured Prediction Model

Structured prediction: predicting a structured output variable \( y \) based on the input variable \( x \).

\( y = (y_1, y_2, \ldots, y_n) \) variables form a structure: sequences, clusters, trees, or arbitrary graphs.

Various approaches have been proposed to learn structured prediction models [Machnies et. al. 09, Chang and Yih 13, Lacoste-Julien et. al. 13] but they are single-threaded.

DEMI-DCD: a multi-threaded algorithm for training structural SVM.

Advantages:

- Requires little synchronization between threads.
- Fully utilizes the power of multiple cores.
- Makes multiple updates on the structures discovered by the loss-augmented inference.
- Fully utilizing the available information.

Inference

\[ \arg \max_{y \in Y} \phi(x, y) \]

Weight parameters.

Set of allowed structures often specified by constraints.

Features on input-output.

Efficient inference algorithms have been proposed for some specific structures.

Integer linear programming (ILP) solver can deal with general structures.

Abstract

Many problems can be framed as structured prediction problems. Structural support vector machines (SVM) is a popular approach for training structured predictors. In structured SVM learning, inference is done by alternating between an inference (prediction) phase and a model update phase. The inference phase selects candidate structures for all training examples and then the model is updated based on these structures. This paper develops an efficient multi-core implementation for structural SVM. We extend the dual coordinate descent approach by decoupling the model update and inference phases into different threads. We prove that our algorithm not only converges but also fully utilizes all available processors to speed up learning.

Motivation:

Inference

Output

Input

Efficient

Structured

Motivation:

Many applications require structured decisions. Global decisions where local decisions play a role but there are mutual dependencies on their outcome. It is essential to make coherent decisions in a way that takes into account the interdependencies among different factors.

Part-of-Speech tagging (sequential labeling)

Input: a sequence of words \( x_1, x_2, \ldots, x_n \).

Output: POS tags \( y_1, y_2, \ldots, y_n \), \( y_i \in \{NN, VBZ, \ldots\} \)

“A cat chases a mouse” => “DT NN VBZ DT NN”.

Assignment to \( y_i \) can depend on both \( x_i \) and \( y_{i-1} \).

Feature vector \( \phi(x, y) \) defined on both input and output variables: e.g., “x; Cat; “, “y1: VBZ”.

Structured Prediction Model

Structured prediction: predicting a structured output variable \( y \) based on the input variable \( x \).

\( y = (y_1, y_2, \ldots, y_n) \) variables form a structure: sequences, clusters, trees, or arbitrary graphs.

Various approaches have been proposed to learn structured prediction models [Machnies et. al. 09, Chang and Yih 13, Lacoste-Julien et. al. 13] but they are single-threaded.

DEMI-DCD: a multi-threaded algorithm for training structural SVM.

Advantages:

- Requires little synchronization between threads.
- Fully utilizes the power of multiple cores.
- Makes multiple updates on the structures discovered by the loss-augmented inference.
- Fully utilizing the available information.

Inference

\[ \arg \max_{y \in Y} \phi(x, y) \]

Weight parameters.

Set of allowed structures often specified by constraints.

Features on input-output.

Efficient inference algorithms have been proposed for some specific structures.

Integer linear programming (ILP) solver can deal with general structures.

Abstract

Many problems can be framed as structured prediction problems. Structural support vector machines (SVM) is a popular approach for training structured predictors. In structured SVM learning, inference is done by alternating between an inference (prediction) phase and a model update phase. The inference phase selects candidate structures for all training examples and then the model is updated based on these structures. This paper develops an efficient multi-core implementation for structural SVM. We extend the dual coordinate descent approach by decoupling the model update and inference phases into different threads. We prove that our algorithm not only converges but also fully utilizes all available processors to speed up learning.