Densified Winner Take All (WTA) Hashing for Sparse Datasets

Beidi Chen
Computer Science Dept.
Rice University
beidi.chen@rice.edu

Anshumali Shrivastava
Computer Science Dept.
Rice University
anshumali@rice.edu

1 PROOF FOR THEOREM 2

Proof: The proof is a simple case based analysis. Define the index of the maximum attribute of vector x as $IndMax(x)$.

$$I_{empty} = \begin{cases} 1 & \Theta(x_1) = \Theta(x_2) = E \\ 0 & otherwise \end{cases} \tag{1}$$

Case I: ($I^i_{empty} = 0$)

Without loss of generality, let $\Theta(x_1)_i \neq E$. If we have $\Theta_i(x_2) \neq E$, then both of these values are untouched and we get

$$\begin{aligned} &\mathcal{H}_{Dwta}(x_1) = \mathcal{H}_{Dwta}(x_2) \\ &\iff IndMax(\Theta_i(x_1)) = IndMax(\Theta_i(x_2)). \end{aligned} \tag{2}$$

In case if $\Theta_i(x_2) = E$, then by the choice of C ,

$$\mathcal{H}_{Dwta}(x_2) > C > IndMax(\Theta_i(x_1)) = \mathcal{H}_{Dwta}(x_1). \tag{3}$$

Therefore, either way we have Equation 2 and

$$\begin{aligned} &Pr(\mathcal{H}_{Dwta}(x_1) = \mathcal{H}_{Dwta}(x_2))|I_{empty} = 0) \\ &= Pr(IndMax(\Theta_i(x_1)) = IndMax(\Theta_i(x_2)))|I_{empty} = 0) \\ &= k_{good}(x1, x2) \end{aligned} \tag{4}$$

Case II: ($I^i_{empty} = 1$)

Let

$$h_u(i, attempt_1)_1 \rightarrow \quad \text{s.t} \quad \Theta_{h_u(i,attempt_1)_1} \neq E.$$

$$h_u(i, attempt_2)_2 \rightarrow \quad \text{s.t} \quad \Theta_{h_u(i,attempt_2)_2} \neq E. \tag{5}$$

$$m = \min(h_u(i, attempt_1)_1, h_u(i, attempt_2)_2). \tag{6}$$

The definition of $h_u(i, attempt_1)_1$ and $h_u(i, attempt_2)_2$ implies $I^m_{empty} = 0$.

Subcase I: ($h_u(i, attempt_1)_1 = h_u(i, attempt_2)_2 = m$)

We have

$$\begin{aligned} &\mathcal{H}_{Dwta}(x_1) = IndMax(\Theta_t(x_1)) + attempt_1 * C \quad \text{and} \\ &\mathcal{H}_{Dwta}(x_2) = IndMax(\Theta_t(x_2)) + attempt_2 * C. \end{aligned} \tag{7}$$

and Equation 2.

Subcase II: ($h_u(i, attempt_1)_1 \neq h_u(i, attempt_2)_2$)
Without loss of generality, $attempt_1 > attempt_2$. Clearly, by definition of m, $h_u(i, attempt)_1$ and $h_u(i, attempt)_2$, we have

$$\begin{aligned} \mathcal{H}_{Dwta}(x_1) &= IndMax(\Theta_{h_u(i,attempt_1)_1)} + attempt_1 * C \\ &> \Theta_{h_u(i,attempt_2)_2} + attempt_2 * C \\ &= \mathcal{H}_{Dwta}(x_2). \end{aligned} \tag{8}$$

Thus, from the two subcases we can write,

$$\begin{aligned} &Pr(\mathcal{H}_{Dwta}(x_1) = \mathcal{H}_{Dwta}(x_2))|I_{empty} = 1) \\ &= Pr(IndMax(\Theta_i(x_1)) = IndMax(\Theta_i(x_2)))|I_{empty} = 0) \\ &= k_{good}(x1, x2) \end{aligned} \tag{9}$$

2 SUPPLEMENTARY PLOTS

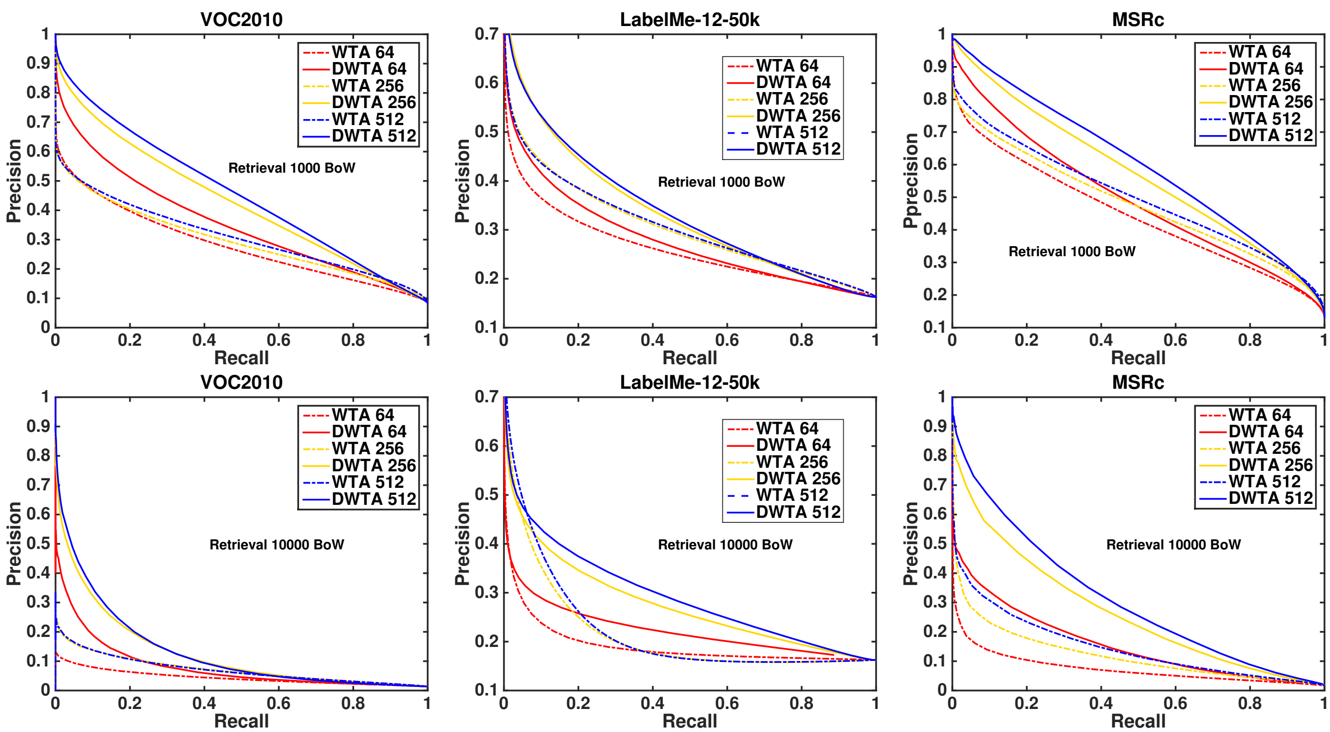

Figure 1: Precision and Recall curves comparing the retrieval performance of Densified WTA vs. WTA on VOC2010, LabelMe-12-50k and MSRc datasets for 1000, 5000 and 10000 BoW feature representations. The semi-dotted lines are the vanilla WTA hashes and bold lines are our proposed Densified WTA Hashes. Different colors represent different number of hashes. We show 1000, 5000, 10000 BoW with 64, 256 and 512 hashes (number of hashes used for ranking) here. Densified WTA significantly outperforms the corresponding WTA consistently.