# The Supplementary Material of the Paper Entitled *Learning Hypergraph-regularized Attribute Predictors*


Sheng Huang
Chongqing University
Chongqing China
huangsheng@cqu.edu.cn

Mohamed Elhoseiny
Rutgers University
Piscataway, NJ, USA
m.elhoseiny@cs.rutgers.edu

Ahmed Elgammal
Rutgers University
Piscataway, NJ, USA
elgammal@cs.rutgers.edu

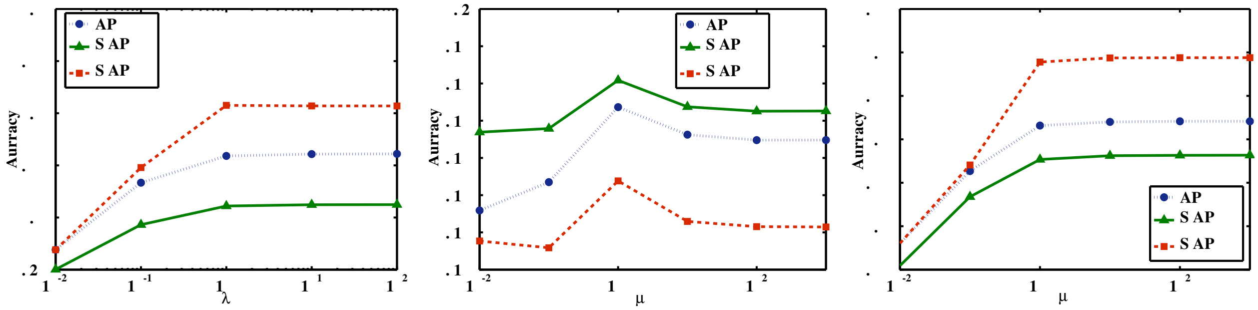
Dan Yang
Chongqing University
Chongqing China
dyang@cqu.edu.cn


## 1. The Derivation Details of Attribute Relation Loss Function

In this section, we will introduce the detailed derivations of Equation (6). Before it, let us review some related notations. The attribute relations are encoded in a given hypergraph $G = (V, E)$ where $V$ is the vertex set and $E$ denotes the hyperedge sets. In this hypegraph, each vertex is corresponding to an instance and each hyperedge is associated with an attribute relation. The degree of a hyperedge $e \in E$, which is denoted as $\delta(e)$, is the number of vertices in $e$. The $(v, e)$-th element of vertex-edge incidence matrix $H \in \mathcal{R}^{|V|\times|E|}$ is considered as $h(v, e) = 1$ if $v \in e$ otherwise $h(v, e) = 0$ where $v \in V$. $d(v) = \sum_{v\in e, e\in E} w(e) = \sum_{e\in E} w(e)h(v, e)$ denotes the degree of the vertex $v$. $w(e)$ is the weight of the hyperedge $e$. We denote the diagonal matrix forms of $\delta(e)$, $d(v)$ and $w(e)$ as $D_e$, $D_v$ and $W$ respectively. We obtain the attribute predictions by defining a collection of hypergraph cuts which is denoted as $F$. $F_u$ returns a row vector of $F$ which is corresponding to the predictions of attributes for the vertex $u$. $L_H$ is the normalized hypergraph Laplacian matrix which is derived from the hypergraph of attributes, and $I$ is an identity matrix. $\mathrm{Tr}(\cdot)$ is the trace of the matrix.

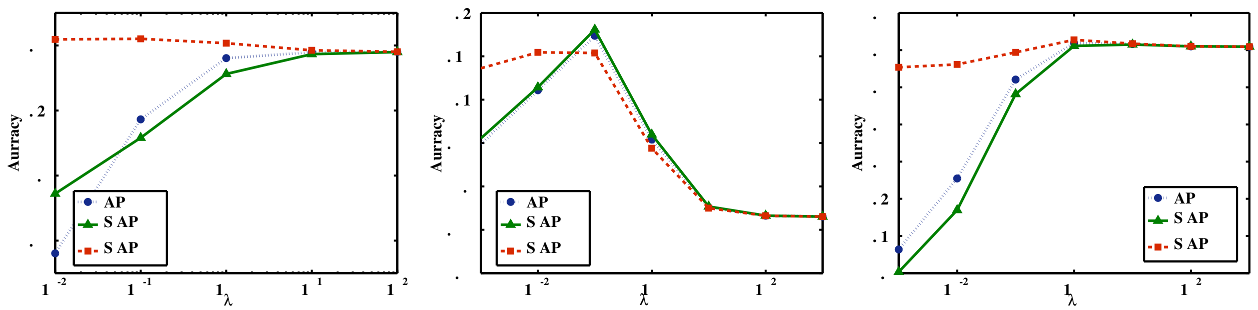
Now we can present the detailed derivations of Equation (6) as follows:

$$\begin{aligned}
\Omega(F, G) &= \frac{1}{2}\sum_{e\in E}\sum_{(u,v)\in e}\frac{w(e)}{\delta(e)}\left\|\frac{F_u}{\sqrt{d(u)}} - \frac{F_v}{\sqrt{d(v)}}\right\|^2 \\
&= \sum_{e\in E}\sum_{u,v\in V}\frac{w(e)h(u,e)h(v,e)}{\delta(e)}(\frac{(F_u)^2}{d(u)} - \frac{F_u F_v^T}{\sqrt{d(u)d(v)}}) \\
&= \sum_{e\in E}\sum_{u\in V}\frac{w(e)h(u,e)(F_u)^2}{d(u)}\sum_{v\in V}\frac{h(v,e)}{\delta(e)} - \sum_{e\in E}\sum_{u,v\in V}\frac{F_u w(e)h(u,e)h(v,e)F_v^T}{\delta(e)\sqrt{d(u)d(v)}} \\
&= \sum_{e\in E}(F_u)^2\sum_{u\in V}\frac{w(e)h(u,e)}{d(u)} - \sum_{e\in E}\sum_{u,v\in V}\frac{F_u w(e)h(u,e)h(v,e)F_v^T}{\delta(e)\sqrt{d(u)d(v)}} \\
&= \sum_{e\in E}(F_u)^2 - \sum_{e\in E}\sum_{u,v\in V}\frac{F_u w(e)h(u,e)h(v,e)F_v^T}{\delta(e)\sqrt{d(u)d(v)}} \\
&= \mathrm{Tr}(F^T F) - \mathrm{Tr}(F^T D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2} F) \\
&= \mathrm{Tr}(F^T (I - D_v^{-1/2} H W D_e^{-1} H^T D_v^{-1/2}) F) \\
&= \mathrm{Tr}(F^T L_H F),
\end{aligned} \quad (1)$$



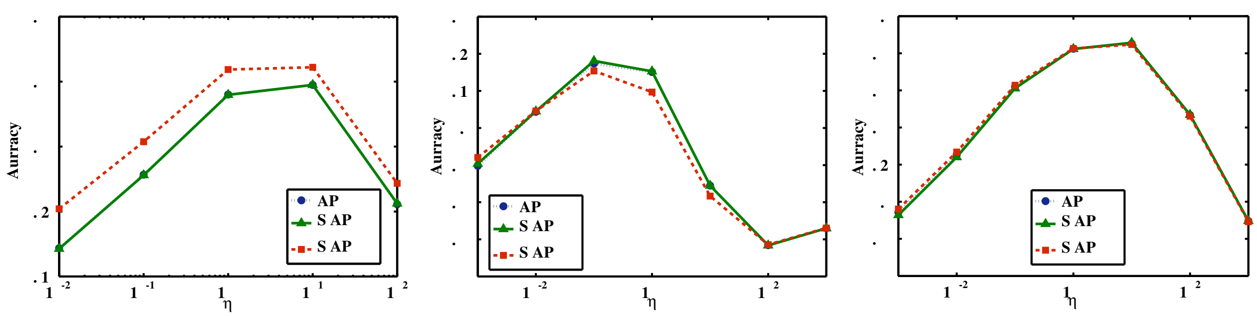

(a) AWA (b) USAA (c) CUB

Figure 1. The influences of $\mu$ to the attribute prediction accuracies.




(a) AWA (b) USAA (c) CUB

Figure 2. The influences of $\lambda$ to the attribute prediction accuracies.

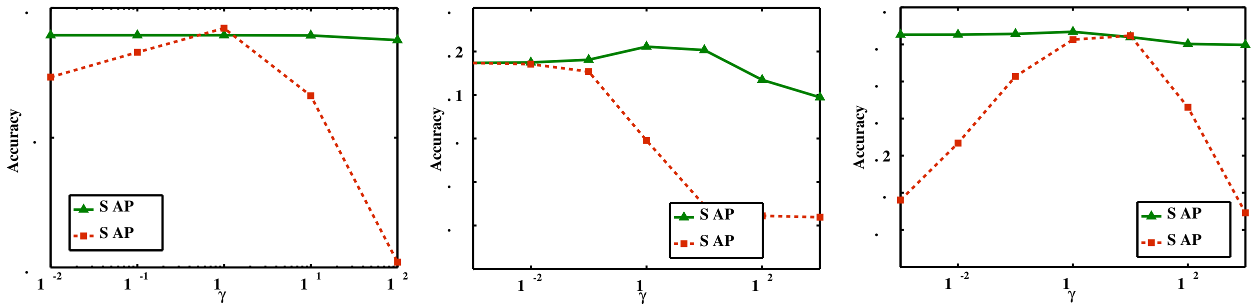



(a) AWA (b) USAA (c) CUB

Figure 3. The influences of $\eta$ to the attribute prediction accuracies.

## 2. The Influences of Parameters

There are several parameters in HAP models. They are $\mu$, $\lambda$, $\eta$ and $\gamma$. $\mu$ is used for controlling the degree of hyperedge weighting. $\lambda$ is used for controlling the trade off between the attribute relation loss and the attribute prediction error. $\eta$ is employed for avoiding the overfitting. $\gamma$ is adopted for controlling the degree of penalty of the side information loss. The ultimate goals of different attribute learning-based systems are different. Some systems may aim at annotation or retrieval. These systems pay more attention on the improvement of the attribute prediction accuracy. Some other systems may focus on the categorization, *i.e.*, Zero-shot Learning, N-shot Learning and Attribute-based categorization. These systems pay more attention on the exploitation of discriminating power of attributes. In our approach, it is available for us to tune the parameters to decide which evaluation metric we care more. Therefore, we will separately discuss the choices of parameters in these two cases.




(a) AWA (b) USAA (c) CUB

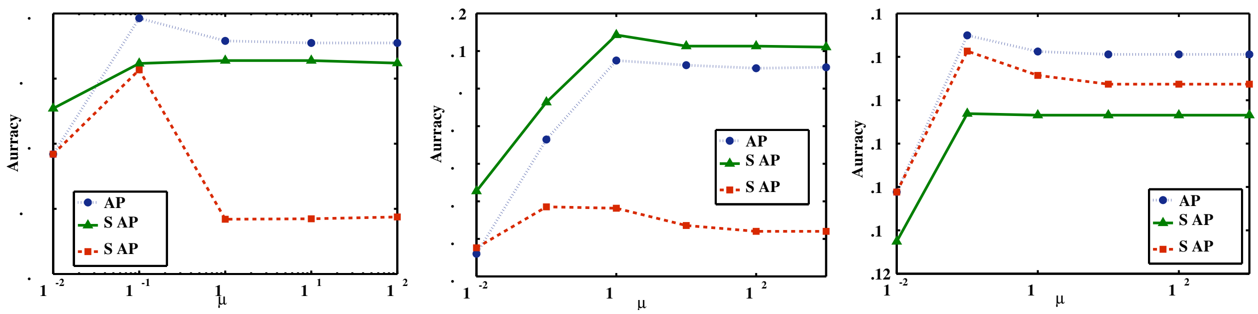
Figure 4. The influences of $\gamma$ to the attribute prediction accuracies.

### 2.1. The Influences of Parameters to The Attribute Prediction

HAP has three parameters, $\mu$, $\lambda$ and $\eta$, need to tuned. CSHAP algorithms have one more parameter $\gamma$ need to be tuned. In the parameter selection procedure, we choose one parameter to tune and fix the values of the other parameters. The initial values of $\mu$, $\lambda$, $\eta$ and $\gamma$ are equal to 0.1. The parameter selection procedure is start from $\mu$ to $\gamma$. Once the optimal value of a parameter is learned, its corresponding initial value is replaced by that optimal value for more accurately estimating the optimal values of the rest parameters. Figures 1, 2 and 3, respectively reports the attribute prediction accuracies using different $\mu$, $\lambda$, $\eta$ on different databases. From the observations, all HAP algorithms can achieve the best performances on all three databases when $\mu = 1$; The best choices of $\lambda$ for AWA, USSA and CUB databases are 10, 0.1 and 1 respectively and such numbers of $\eta$ are 10, 0.1 and 10. Figure 4 plots the relationships between $\gamma$ and the attribute prediction accuracy on three different databases. We can find that CSHAP$_G$ is more sensitive to $\gamma$. It is not hard to conclude from the observations that the optimal values of $\gamma$ are 1, 0.01 and 10 for AWA, USAA and CUB databases respectively.

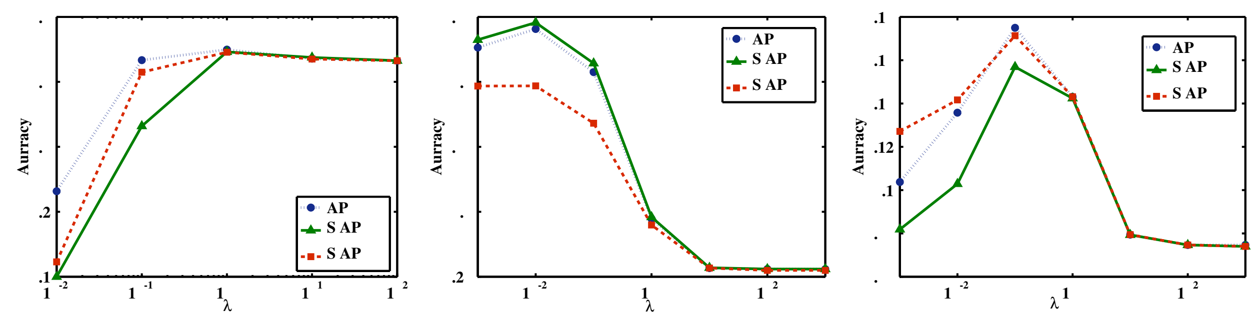
### 2.2. The Influences of Parameters to Zero-Shot Learning

In Zero-Shot Learning (ZSL), we need to employ the sigmoid function to normalize the attribute confidences, which are obtained by our models, into range [0,1]. So, there is one additional parameter $\rho$ should be studied in this section. We follow the aforementioned parameter selection manner to select the parameters. The selection procedure is start from $\mu$ to $\rho$ where the initial value of $\rho$ is 0.5. Figure 5 shows the ZSL accuracies under different $\mu$. On AWA and CUB database, all three approach can get the best performances when $\mu = 0.1$ while the optimal value of $\mu$ on USAA database is 1. Compared with other parameters, HAP algorithms are relatively insensitive to $\mu$ when its value is bigger than 1. Figures 6 and 7 demonstrate the impacts of ZSL accuracies from $\lambda$ and $\eta$ respectively. The curves of these figures share similar behavior that their peaks are very explicit. From the observations, we can know that the optimal values of $\lambda$ are 1, 0.01 and 0.1 on AWA, USAA and CUB databases respectively while such numbers of $\eta$ are 1, 0.1 and 0.1. As same as the phenomenon observed in Figure 4, Figure 8 also shows that CSHAP$_G$ is very sensitive to $\gamma$ but CSHAP$_H$ is robust to $\gamma$. Here, we suggest to set the $\gamma$ of AWA, USAA and CUB databases to 1, $10^{-3}$ and $10^{-3}$ respectively. Figure 9 reports the ZSL performances under different $\rho$. However, it is really hard to conclude uniform setting for each database. So we choose different $\rho$ for different approaches. More specifically, we suggest to choose the $\rho$ in the range [0.007, 0.02] for CSHAP$_H$ while choose the $\rho$ in the range [0.06,0.1] for HAP and CSHAP$_G$ on AWA database. On USAA database, CSHAP$_G$ can get good ZSL performances when $\rho$ is in the range [0.005,0.01] while the good $\rho$ for CSHAP$_H$ and HAP should be above 0.3. The impacts of $\rho$ to the performances of all three algorithms are similar on CUB databases. The observations indicate that the $\rho$ which is larger than 1 can get the good performances for all three HAP algorithms.



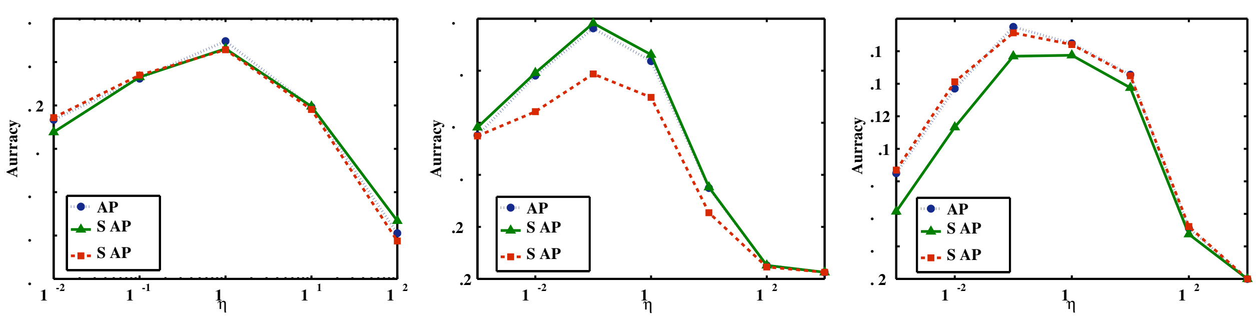

(a) AWA (b) USAA (c) CUB

Figure 5. The influences of $\mu$ to the ZSL accuracies.



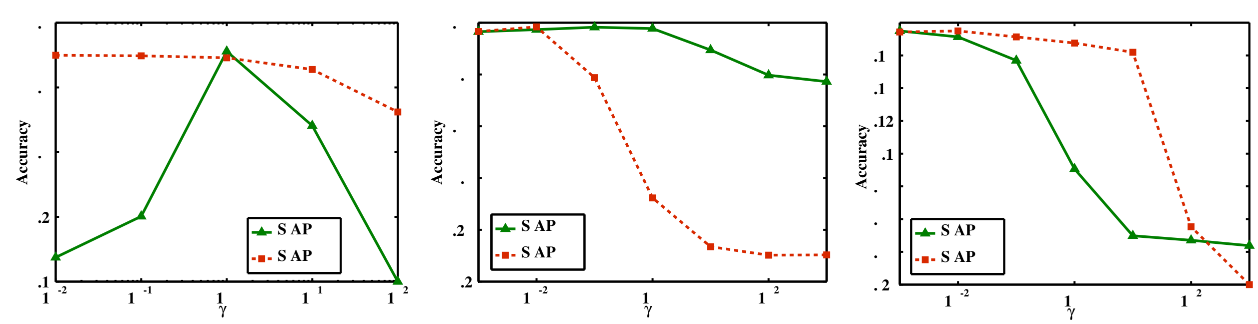

(a) AWA (b) USAA (c) CUB

Figure 6. The influences of $\lambda$ to the ZSL accuracies.



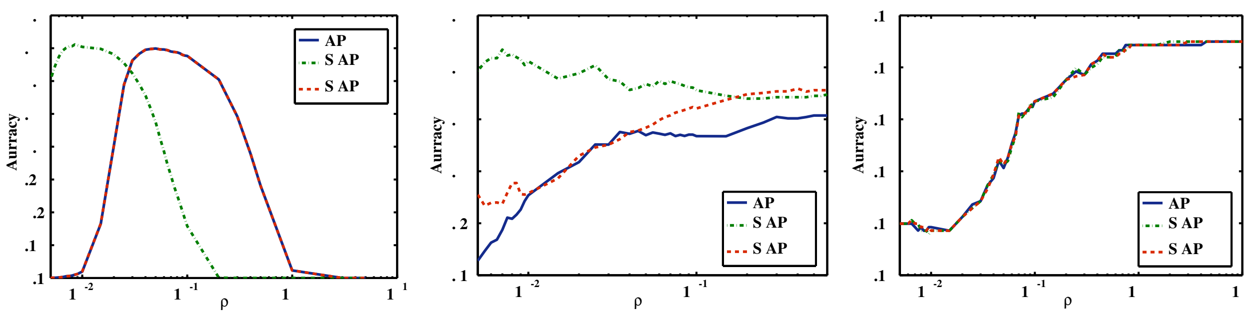

(a) AWA (b) USAA (c) CUB

Figure 7. The influences of $\eta$ to the ZSL accuracies.




(a) AWA (b) USAA (c) CUB

Figure 8. The influences of $\gamma$ to the ZSL accuracies.




(a) AWA (b) USAA (c) CUB

Figure 9. The influences of $\rho$ to the ZSL accuracies.