Proof of Propositions

**Proposition 6.1.** For any kernel \( \mathbf{w} \) of size \( k \), for all \( n \geq \left(\frac{5k - 1}{2}\right) \) if \( k \) is odd and \( n \geq \left(\frac{(6k+2)(k-1)}{3k}\right) \) if \( k \) is even, the following statement holds:

\[
\mathcal{N}_{het}(\phi_{alt},k) \geq \mathcal{N}_{het}(\phi_{stk},k)
\]

**Proof.** Any \( M_k \in \mathbb{R}^{k \times k}, \ M_k \subseteq \phi_{alt} \) contains \( \left\lfloor \frac{k}{2} \right\rfloor \) rows of elements of \( e_s \), and \( \left\lfloor \frac{k+1}{2} \right\rfloor \) rows of elements of \( e_r \), or vice-versa.

For a single fixed \( M_k \), the total number of triples \((a_i, b_j, M_k)\) and \((b_j, a_i, M_k)\) is

\[
2 \times k \left\lfloor \frac{k}{2} \right\rfloor \times k \left\lfloor \frac{k+1}{2} \right\rfloor
\]

The number of possible \( M_k \) matrices is \((n-k+1)^2\). Hence the total number of heterogeneous interactions is

\[
\mathcal{N}_{het}(\phi_{alt},k) = (n-k+1)^2 \times 2 \left\lfloor \frac{k}{2} \right\rfloor \left\lfloor \frac{k+1}{2} \right\rfloor
\]

Any \( M_k \in \mathbb{R}^{k \times k}, \ M_k \subseteq \phi_{stk} \) contains \( l \) rows of elements of \( e_s \), and \( k - l \) rows of elements of \( e_r \), where \( 0 \leq l \leq k \).

For a fixed \( l \), the number of different possible \( M_k \) matrices is \((n-k+1)\). Hence, the number of heterogeneous interactions is

\[
\mathcal{N}_{het}(\phi_{stk},k) = (n-k+1) \left( \sum_{l=0}^{k} 2 \times kl \times k(k-l) \right)
\]

\[
= (n-k+1) \cdot k^2 \cdot \left( k^2(k+1) - \frac{k(k+1)(2k+1)}{3} \right)
\]

\[
= (n-k+1) \cdot k^2 \cdot \left( \frac{k(k+1)(k-1)}{3} \right)
\]

(1)

We need to check whether,

\[
\mathcal{N}_{het}(\phi_{alt},k) \geq \mathcal{N}_{het}(\phi_{stk},k)
\]

For odd \( k \), this becomes

\[
(n-k+1) \left( \frac{k-1}{2} \right) \left( \frac{k+1}{2} \right) \geq \frac{k(k+1)(k-1)}{6}
\]

\[
\frac{k-1}{2} + 1 \geq \frac{k+1}{2}
\]

\[
n - k + 1 \geq \frac{2k}{3}
\]

\[
n \geq \frac{5k}{3} - 1
\]

For even \( k \),

\[
(n-k+1) \left( \frac{k}{2} \right)^2 \geq \frac{k(k+1)(k-1)}{6}
\]

\[
nk - k(k-1) \geq \frac{2(k+1)(k-1)}{3}
\]

\[
n \geq \frac{(5k+2)(k-1)}{3k}
\]

**Proposition 6.2.** For any kernel \( \mathbf{w} \) of size \( k \) and for all \( \tau < \tau' \ (\tau, \tau' \in \mathbb{N}) \), the following statement holds:

\[
\mathcal{N}_{het}(\phi_{alt}^{\tau},k) \geq \mathcal{N}_{het}(\phi_{alt}^{\tau'},k)
\]

**Proof.** For simplicity, let us assume that \( \alpha = n/(2\tau) \in \mathbb{N} \), i.e., \( \phi_{alt}^{\tau} \) is composed of exactly \( \alpha \) blocks of \( \tau \) rows of \( e_s \) and \( e_r \) stacked alternately. Also, when \( \tau < k \), we assume that \( k/\tau \in \mathbb{N} \). Now, for any \( M_k \in \mathbb{R}^{k \times k}, \ M_k \subseteq \phi_{alt}^{\tau}, \) we consider the following two cases:

**Case 1.** \( \tau \geq k - 1 \): It is easy to see that this case can be split into \( n/\tau \) subproblems, each of which is similar to \( \phi_{stk}^{\tau} \). Hence,

\[
\mathcal{N}_{het}(\phi_{alt}^{\tau},k) = \left( \frac{n}{\tau} \right) \mathcal{N}_{het}(\phi_{stk},k)
\]

Clearly, \( \mathcal{N}_{het}(\phi_{alt}^{\tau},k) \) is monotonically decreasing with increasing \( \tau \).
Case 2. $\tau < k - 1$: As shown in Fig. 1 let $T_a, T_b \in \mathbb{R}^{\tau \times k}$ denote a submatrix formed by components of $e_s, e_r$ respectively. Note that if $k$ is even, then for any $M_k \subseteq \Phi_{alt}$, the number of components of $e_s$ and $e_r$ are always equal to $k^2/2$ each. For odd $k$, the number of $T_a$’s and $T_b$’s are $(k/\tau+1)$ and $(k/\tau-1)$ in some order. Now, if we move $M_k$ down by $i$ rows ($i \leq \tau$), the total number of heterogeneous interactions across all such positions is:

$$
\begin{align*}
\sum_{i=0}^{\frac{k^2}{\tau}} \left(\frac{k}{\tau}i + \frac{1}{2}\right) \tau - \left(\frac{k}{\tau}i - \frac{1}{2}\right) \tau + i \\
= \frac{n k^2}{\tau} \sum_{i=0}^{k-1} (k + \tau - 2i)(k - \tau + 2i) \\
= \frac{n k^2}{4} \sum_{i=0}^{k-1} k^2 - (\tau^2 + 4i^2 - 4i\tau) \\
= \frac{n k^2}{4} \left((k^2 - \tau^2) - \frac{4(\tau - 1)(2\tau - 1)}{6} + \frac{4\tau(\tau - 1)}{2}\right) \\
= C \left(k^2 - \frac{\tau^2}{3} + \frac{2}{3}\right)
\end{align*}
$$

We can see that this is also monotonically decreasing with increasing $\tau$. It is also evident that the above expression is maximum at $\tau = 1$ (since $\tau \in \mathbb{N}$).

Proposition 6.3. For any kernel $w$ of size $k$ and for all reshaping functions $\phi : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^{n \times n}$, the following statement holds:

$$
\mathcal{N}_{het}(\phi_{chk}, k) \geq \mathcal{N}_{het}(\phi, k)
$$

Proof. For any $\phi$ and for any $M_k \in \mathbb{R}^{k \times k}$ such that $M_k \subseteq \phi$, let $x, y$ be the number of components of $e_s$ and $e_r$ in $M_k$ respectively. Then $\mathcal{N}_{het}(M_k, k) = 2xy$. Also, since total number of elements in $M_k$ is fixed, we have $x + y = k^2$.

Using the AM-GM inequality on $x, y$ we have,

$$
xy \leq \left(\frac{x + y}{2}\right)^2 = \frac{k^4}{4}
$$

If $k$ is odd, since $x, y \in \mathbb{N}$,

$$
xy \leq \frac{k^4 - 1}{4}
$$

Therefore, the maximum interaction occurs when $x = y = \frac{k^2}{2}$ (for even $k$), or $x = \frac{k^2}{2} + 1, y = \frac{k^2 - 1}{2}$ (for odd $k$). It can be easily verified that this property holds for all $M_k \subseteq \Phi_{chk}$.

Hence,

$$
\mathcal{N}_{het}(\phi, k) = \sum_{M_k \subseteq \Omega_c} 2xy \leq \sum_{M_k \subseteq \Omega_c} \frac{2k^4}{4} = \mathcal{N}_{het}(\phi_{chk}, k)
$$

Proof. If $M_k$ contains $x$ components of $e_s$ and $y$ components of $e_r$, then $\mathcal{N}_{het}(M_k, k) = 2xy$, and $\mathcal{N}_{het}(M_k, k) = 2(x - l)(y - (p - l))$ for some $l \leq p$ and $l \leq x$. We observe that

$$
\mathcal{N}_{het}(M_k', k) = \mathcal{N}_{het}(M_k, k) - 2(x - l)(p - l) - 2ly \leq \mathcal{N}_{het}(M_k, k)
$$

Proposition 6.4. Let $\Omega_0, \Omega_c : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{(n+p) \times (n+p)}$ denote zero padding and circular padding functions respectively, for some $p > 0$. Then for any reshaping function $\phi$,

$$
\mathcal{N}_{het}(\Omega_c(\phi), k) \geq \mathcal{N}_{het}(\Omega_0(\phi), k)
$$

Proof. Given $\Omega_c(\phi)$, we know that we can obtain $\Omega_0(\phi)$ by replacing certain components of $\Omega_c(\phi)$ with $0$. So for every $M_k \subseteq \Omega_c(\phi)$, there is a corresponding $M_k' \subseteq \Omega_0(\phi)$ which is obtained by replacing some $p$ components ($p \geq 0$) of $M_k$ with $0$.

Using the above Lemma, we can see that

$$
\begin{align*}
\mathcal{N}_{het}(\Omega_c(\phi), k) &= \sum_{M_k \subseteq \Omega_c(\phi)} \mathcal{N}_{het}(M_k, k) \\
&\geq \sum_{M_k \subseteq \Omega_0(\phi)} \mathcal{N}_{het}(M_k', k) \\
&= \mathcal{N}_{het}(\Omega_0(\phi), k)
\end{align*}
$$

Hyperparameters

We use the standard training, validation and test splits provided with the datasets. A detailed description of the datasets is included in the main paper. We select the best model using the validation data on the hyperparameters listed in Table 1. Most of the hyperparameters are adopted from ConvE (Dettmers et al. 2018) model. In this paper, we explore both 1-1 (Bordes et al. 2013) and 1-N (Dettmers et al. 2018) scoring techniques. In 1-N scoring, each $(s, r)$ pair is scored against all the entities $o \in E$ simultaneously. For training, we use Adam optimizer (Kingma and Ba 2014) and Xavier initialization (Glorot and Bengio 2010) for initializing parameters.
Hyperparameter Values

<table>
<thead>
<tr>
<th>Hyperparameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate</td>
<td>[0.001, 0.0001]</td>
</tr>
<tr>
<td>Label smoothing</td>
<td>[0.0, 0.1]</td>
</tr>
<tr>
<td>Batch size</td>
<td>{16, 64, 256}</td>
</tr>
<tr>
<td>Negative Samples</td>
<td>{100, 500, 1000, 4000}</td>
</tr>
<tr>
<td>$l_2$ regularization</td>
<td>[0, $10^{-5}$]</td>
</tr>
<tr>
<td>Hidden dropout</td>
<td>[0, 0.3, 0.5]</td>
</tr>
<tr>
<td>Feature dropout</td>
<td>[0, 0.2, 0.5]</td>
</tr>
<tr>
<td>Input dropout</td>
<td>[0, 0.2]</td>
</tr>
<tr>
<td>$k_w$</td>
<td>[10]</td>
</tr>
<tr>
<td>$k_h$</td>
<td>[20]</td>
</tr>
<tr>
<td>Number of convolutional filters</td>
<td>{32, 48, 64, 96}</td>
</tr>
<tr>
<td>Convolutional kernal size $k$</td>
<td>[3, 5, 7, 9, 11]</td>
</tr>
<tr>
<td>Number of feature permutations $t$</td>
<td>{1, 2, 3, 4, 5}</td>
</tr>
</tbody>
</table>

Table 1: Details of hyperparameters used. Please refer to Section for more details.

References


