

# Gyógyszer kölcsönhatások azonosítása gráf autoenkóderrel

Bilibók Bence, Péter István, Rancz Máté

# A téma motivációja

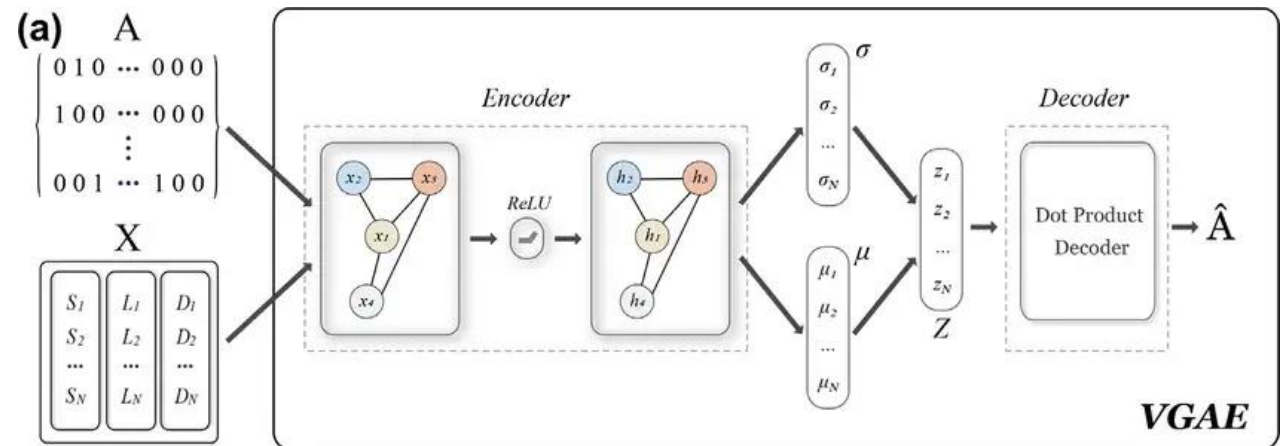
- A gyógyszeriparban egy hatóanyag biztonságossága szempontjából fontos, hogy milyen más hatóanyagokkal lép kölcsönhatásba
- Ezek az *adverz* reakciók, legyen erősítő vagy gyengítő
- Többnyire kísérleti módszerek
- DL alapú predikció

# Probléma megfogalmazása

- Élpredikció gráfon
- A csúcsok a hatóanyagok
- Az élek adverz reakciók, ha vannak

# Modell: VGAE

- Variációs Gráf Autoenkóder
- Gráf konvolúciós réteg
- Látens reprezentáció
- Mintavételezés egy normál eloszlásból
- Két csúcs látens címkei közti viszony határozza meg, hogy össze lehetnek-e kötve
- Skaláris szorzat



# Modell: VGAE

**Inference model** We take a simple inference model parameterized by a two-layer GCN:

$$q(\mathbf{Z} | \mathbf{X}, \mathbf{A}) = \prod_{i=1}^N q(\mathbf{z}_i | \mathbf{X}, \mathbf{A}), \quad \text{with} \quad q(\mathbf{z}_i | \mathbf{X}, \mathbf{A}) = \mathcal{N}(\mathbf{z}_i | \boldsymbol{\mu}_i, \text{diag}(\boldsymbol{\sigma}_i^2)). \quad (1)$$

Here,  $\boldsymbol{\mu} = \text{GCN}_{\boldsymbol{\mu}}(\mathbf{X}, \mathbf{A})$  is the matrix of mean vectors  $\boldsymbol{\mu}_i$ ; similarly  $\log \boldsymbol{\sigma} = \text{GCN}_{\boldsymbol{\sigma}}(\mathbf{X}, \mathbf{A})$ . The two-layer GCN is defined as  $\text{GCN}(\mathbf{X}, \mathbf{A}) = \tilde{\mathbf{A}} \text{ReLU}(\tilde{\mathbf{A}} \mathbf{X} \mathbf{W}_0) \mathbf{W}_1$ , with weight matrices  $\mathbf{W}_i$ .  $\text{GCN}_{\boldsymbol{\mu}}(\mathbf{X}, \mathbf{A})$  and  $\text{GCN}_{\boldsymbol{\sigma}}(\mathbf{X}, \mathbf{A})$  share first-layer parameters  $\mathbf{W}_0$ .  $\text{ReLU}(\cdot) = \max(0, \cdot)$  and  $\tilde{\mathbf{A}} = \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}}$  is the symmetrically normalized adjacency matrix.

**Generative model** Our generative model is given by an inner product between latent variables:

$$p(\mathbf{A} | \mathbf{Z}) = \prod_{i=1}^N \prod_{j=1}^N p(A_{ij} | \mathbf{z}_i, \mathbf{z}_j), \quad \text{with} \quad p(A_{ij} = 1 | \mathbf{z}_i, \mathbf{z}_j) = \sigma(\mathbf{z}_i^T \mathbf{z}_j), \quad (2)$$

where  $A_{ij}$  are the elements of  $\mathbf{A}$  and  $\sigma(\cdot)$  is the logistic sigmoid function.

**Learning** We optimize the variational lower bound  $\mathcal{L}$  w.r.t. the variational parameters  $\mathbf{W}_i$ :

$$\mathcal{L} = \mathbb{E}_{q(\mathbf{Z} | \mathbf{X}, \mathbf{A})} [\log p(\mathbf{A} | \mathbf{Z})] - \text{KL}[q(\mathbf{Z} | \mathbf{X}, \mathbf{A}) || p(\mathbf{Z})], \quad (3)$$

# Jellemzők kinyerése

- Adatbázis vizsgálata
- Sok jellemző szöveggel van leírva -> nem megfelelő nekünk
- Számszerű jellemzők kiválasztása (tömeg, olvadási pont, stb.)
- Molekuláris összetétel és osztályozások -> one-hot kódolással

# Tanítás

- <https://github.com/dmlc/dgl/tree/master/examples/pytorch/vgae>
- Binary cross entropy (élpredikció) + KL divergence (látens eloszlás)
- Az élek 5 ill. 10%-a teszt és validációs él (maszkolás a tanítás során)
- <https://github.com/peter-i-istvan/nhf-drugs>
- 0.05 learning rate, Adam
- 10 hosszú bemeneti vektor (feature), mint címke
- 32 dimenziós látens tér

# Eredmények, kiértékelés

