

Clasificación tipo de relaciones entre Sistemas Autónomos

Benchmark

Para abordar este problema se comenzó por la creación de un Benchmark.

Para esto se probaron los siguientes métodos anteriores:

- Gao [sacado de BGP2VEC]
- Ruan [sacado de BGP2VEC]
- Sark [sacado de BGP2VEC]
- AS Rank (CAIDA)
- Toposcope

Datos

- Se usó x durante los X días
- ¿Cuánto se demoró?

Modelos

Para abordar este problema se utilizaron

Dataset de Evaluación

Las relaciones entre SA son privadas, esto hace que en algoritmos de inferencia sea difícil validar, y para casos de machine learning, entrenar los modelos.

Lo primero que podríamos hacer es buscar un SA que publique sus AS relaciones. Esto fue lo que hice la persona de la tesis y creo un dataset a partir de la información que provee Hurricane's BGP Toolkit.

Hurricane BGP Toolkit

El Hurricane Electric BGP Toolkit es un conjunto de herramientas en línea desarrollado por Hurricane Electric, diseñado para analizar y monitorear datos de BGP. <https://bgp.he.net/>

Experimentos

Para abordar este problema se tomaron dos enfoques diferentes.

- Clasificación Binaria: Se tomaron p2c y c2p como una misma clase y las otras siendo p2p.
- Clasificación Multiclase: Las relaciones p2c y c2p se tomaron como clases diferentes.

Dataset de Validación

Uno de los posibles caminos a ocupar fue ocupar como dataset de validación de la Tesis "BGP Data Analysis: Exploring Solutions for Autonomous Systems Relationships Inference"

Este utiliza el atributo BGP communities más la información RIB, esto con la idea de que hay algunos AS que publican sus políticas de BGP communities. Entonces se creó un JSON que contiene reglas de ciertos SA y reglas (BGP communities policy) con las que podría saber el tipo de relación que se tendría con un vecino.

Por ejemplo TODO: Agregar un ejemplo.

Experimento 1:

- GNN -> GCN
- Predictor -> DotProduct y MLP
- Optimizador -> Adam

- Función de pérdida -> CrossEntropyLoss
- Split de edges para entrenamiento y validacion
- Stochastic Gradient Descent (SGD) con un learning rate de XXX.



Figure 1: Resultados.

- Problemas con el modelo:
 - Posible overfitting.

Experimento 2:

- GNN -> GraphSAGE
- Predictor -> MLP
- Optimizador -> Adam
- Función de pérdida -> CrossEntropyLoss
- Split de edges para entrenamiento y validacion
- Neighbour sampling.

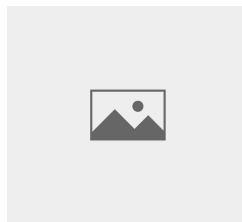


Figure 2: Resultados.

- ¿Por que ocupamos GraphSAGE y no GCN?

Como estamos entrenando de forma transductiva, es decir ocupando un mismo grafo para entrenar y validar, puede ocurrir que se este overfitando el grafo y por eso obteniendo buenos resultados, esto es debido a que al probar con epoch muy grandes los valores de loss todo el rato eran muy similares, y lo que se espera obtener en este caso es que llegase a un punto donde la loss era similar, pero luego empezaran a diverger, que seria el punto en donde el modelo se esta empezando a aprender los datos de memoria, pero esto no paso???. Para evitar esto se decidio ocupar otras formas de entrenamiento para mejorar que el modelo pudiese generalizar y no se estuviese validando mal (porque al final mas que nada es problema de que no estamos validando con datos diferentes). Entrenando con sampling.

Experimento 3:

- GNN -> GraphSAGE
- Predictor -> MLP
- Optimizador -> Adam
- Función de pérdida -> CrossEntropyLoss
- Split de edges para entrenamiento y validacion
- Otro sampling



Figure 3: Resultados.

Experimento 4:

Agregar paquetes de flujo



Figure 4: Resultados.

Destacar que en nuestro enfoque fue de Regresión, es decir la salida final de los modelos corresponde a un valor continuo al cual luego se calcula un umbral para clasificar en las clases correspondientes. Esto porque al tener clases desbalanceadas, el modelo desearía en solo clasificar un solo tipo (la que es mayor) y aun así no se tienen resultados tan malos. Por esto no se tomó ese enfoque ya que es una de las fallas a la que queda propenso.

GCN VS GraphSAGE

- Comparamos el caso de GCN con GraphSAGE usando como Predictor el DotProduct y ocupando todas las features:

GCN:

- Con 100 epoch, un % de train de 0.6 se obtiene una accuracy de 0.81, sin embargo se puede observar overfitting más o menos desde el epoch 70.
- Ocupando un modelo que agrega entre cada capa GNN dropout se obtiene una accuracy 0.8. No es muy diferentes.
- Drop out ayuda mucho, sin embargo hace que no sea tan smooth el converger.

GraphSAGE:

- Con epoch 100, un % de experience

Podemos ver que GraphSAGE super en performance a GCN. También podemos ver que para ambos modelos el overfitting se presenta en diferentes casos, esto por la naturaleza de los cálculos de cada uno en la agregación. Pues en GCN para obtener un mensaje al nodo, se necesitan todos sus vecinos haciendo que sea difícil reconocer un nodo de otro ya sea más fácil de que se reconozca overfitting, en cambio en GraphSAGE el mensaje al nodo es como un promedio de los vecinos, por lo que es más difícil que se reconozca overfitting.

Se realizó la misma comparación pero en vez de DotProduct se ocupó MLP Predictor. En el caso de GCN la accuracy subió a un 0.8222(0.9221 cambiando porte capas) y GraphSAGE a un 0.92 (con el mismo número entre capas) 0.9546 (cambiando número de capas). Sin embargo con 100 epoch no cambiaba mucho al final. Quise ver que ocurría si ocupaba un porcentaje muy bajo de train, para ver si había overfitting, pero lo que pasó fue que igual había buenos resultados...??? ocupé el modelo sin dropout y ahí se podía ver un poco el overfitting. con una cantidad de X ejemplos de edge para train.

[Cachar porque no me esta encajando el numero de true y falses en el train msakk al ahcer split del dataset]

como sea Es mejro con MLP que con DotProduct, por lo que se ocupara MLP en los siguientes experimentos.

OCUPAR TODAS LAS FEATURES ES NECESARIO?

Luego para ver que tan esenciales en la tarea era los features recolectados se porbo con graphSAGE y MLP como predictor (con ahsta el momento los mejores resultados que s e ahan obtenido), con 100 epoch y un % de train de 0,6 .

Se partio ocupando unicamente como atributos de los ndoos su grado in y grado out por nodo. Obteniendo una Accuracy de 0.8724 con DotProduct y 0.9290 con MLPpredcitor. Con esto vemos que si bien se dan buenos rsulttodos agregar las feeatures Mejora la accuracy. ¿Pero son de ayuda las que le estamos pasando? Prque si bien ya tenemos que son una ayuda luego de hacer una exploracion de esta (En anexo mas info) podemos notar que para muchos features par ala mayria de los nodos no se tiene información y por ende no serian muy relevantes?.

PAra esto se decidio incluir unicamete aquellas features cuya información para todos los nodos estuviera sobre 80%. Es decir existe info de la feature para el 80 % de los nodos. Estoso consistieron en :

- AS_rank_numberAsns
- AS_rank_customer
- AS_rank_peer
- peeringDB_ix_count
- peeringDB_fac_count
- cti_top

Con estos usando GraphSAGE y MLP como predictor se obtuvo una accuracy de 0.9452 lo que s eve que tener todas las otras valores lo mejora pero no tanto, es decir no son tan relevantes para dicha tarea.

otros caso fue elegir unicamente[COMPLETAR]...

IDEAS FALLIDAS

- Dentrro de otras ideas que se intentaron pero no funcionaron fue crear grafos a partir de los recolectores RRC, de sta forma tener diferentes grafos a partri d elos cuales algunos se elijan para training y otro diferente para testeo. Sin embargo falle, no se si retomar o no.
- otra idea fue que RIPEstat tineen una API de la cual se puede obtener información, sin embrago se demora demasiado para la cantidad de nodos ques e tienen por grafos.

Continuando con los experimentos una de los problemas que mas miedo tenia era que se entrenara y se esttuviera overfitteoando el grafo, porque no tenemso mas grafo que el de esa fecha, pues esos datos (los atributos corresponden a sacados por otro paper) . Es decir estabamos tomando un enfoque inductivo. Es por esto que una podible solución que se nos ocurrio fue proabr diferentes tecnicas de samplingm, area que aun sigue en investigación e inovacion en el area de GNN.

Para esto se partio con random neighbour sampling[explicado en XXX marco teorico] y se obtuvo una accuracy de 0.9414 , lo que no mejoro mucho los resultados obtenidos anteriormente. ya desde el segundo epoch hay overfitiing.

se decidió probar con otro tipo de método el cual correspondiente a ClusterGCN [explicado marco teórico sección XXX] obteniendo una accuracy 0.9696 !!!!!!!

Luego para comparar GNN con otros métodos para resolver dicha tarea, se probó resolver la tarea de clasificación con PageRank, DeepWalk y BGP2Vec. Obteniendo Resultados XX, 0.9270 (entrenando) y XX respectivamente.

La idea es ver que tan bien se comporta el modelo en comparación con otros métodos que se han ocupado para resolver la misma tarea.

Caso	Accuracy
GCN + DotProductPredictor	0.81
GraphSAGE + DotProductPredictor	0.9
GCN + MLPredictor	0.9221
GraphSAGE + MLPredictor	0.9546
GraphSAGE + MLPredictor + in_degree y out_degree	0.9290
GraphSAGE + MLPredictor + features sobre 80%	0.9452
GraphSAGE + MLPredictor + Random neighbour sampling	0.9414
GraphSAGE + MLPredictor + ClusterGCN	0.9696
PageRank	0.8746
DeepWalk	0.9270
BGP2Vec	X
Crear de 0	0.9068

Con GCN el overfitting es más fácil que con GraphSAGE, es de esperarse por la fórmula de cada una. GraphSAGE para que se vea overfitting % de train debe ser más bajo a diferencia de GCN que puede ser alto .