

بررسی ویژگی‌های گراف در یادگیری ماشین

سجاد سبزی
محمدرضا احمدی تشنیزی

۲۷ آبان ۱۴۰۲

مقدمه

ویژگی‌های گراف در یادگیری ماشین یک حوزه پیچیده و حیاتی برای مطالعه می‌باشند. آنها از ساختارهای گراف برای مدل‌سازی روابط و تعاملات پیچیده در انواع مختلف داده‌ها بهره می‌برند. این روش در زمینه‌های متنوعی کاربرد دارد و از تجزیه و تحلیل شبکه‌های اجتماعی تا تفسیر داده‌های بیولوژیکی متنوعی پیشرفته استفاده می‌شود. بخش‌های بعدی به بررسی جامع اصول اساسی و کاربردهای آنها در زمینه یادگیری ماشین می‌پردازند.

ساختار و نمایش گراف

گراف‌ها که از گره‌ها و یال‌ها تشکیل شده‌اند، عناصر اساسی برای نمایش موجودیت‌ها و روابط آنها می‌باشند. گراف‌های جهت‌دار جهت روابط را تمایز می‌دهند، در حالی که گراف‌های بدون جهت هیچ جهت خاصی را نمایان نمی‌کنند. به علاوه، گراف‌های وزن‌دار مقداری به یال‌ها اختصاص می‌دهند که قدرت یا ظرفیت اتصالات را نمایان می‌کنند.

انواع گراف‌ها در یادگیری ماشین

- شبکه‌های اجتماعی: این‌ها ساختارهای اجتماعی را نمایان می‌کنند، که گره‌ها را به عنوان افراد و یال‌ها را به عنوان تعاملات اجتماعی می‌پذیرند.
- شبکه‌های بیولوژیکی: این موارد شامل شبکه‌هایی مانند تعاملات پروتئین-پروتئین و چارچوب‌های تنظیم ژنتیکی می‌شوند.
- گراف‌های دانش: به طور اصلی در وب معنایی و پردازش زبان طبیعی استفاده می‌شوند و روابط میان داده‌ها را نمایش می‌دهند.

تکنیک‌های تجزیه و تحلیل گراف

- اندازه‌گیری مرکزیت: تکنیک‌هایی مانند درجه، نزدیکی، بینایی و مرکزیت مقادیر بحرانی را در یک شبکه شناسایی می‌کنند.
- شناسایی انجمن: این شامل شناسایی خوشه‌های گره که ارتباطات درونی آنها نسبت به باقی گراف چگونه است را می‌پذیرد.
- تجزیه و تحلیل جریان شبکه: در گراف‌های وزن‌دار و جهت‌دار برای مدل‌سازی پدیده‌هایی مانند جریان ترافیک و انتشار اطلاعات به کار می‌رود.

مدل‌های یادگیری ماشین مبتنی بر گراف

- شبکه‌های عصبی گراف (GNNs) این شبکه‌ها برای پردازش داده‌های گرافی ساخته شده‌اند و وابستگی‌های میان گره‌ها را ضبط می‌کنند.
- شبکه‌های تبدیلی گراف (GCNs) این یک زیرنوع از GNN هستند که از روش تبدیلی برای مدیریت داده‌های گرافی استفاده می‌کنند.
- تعبیرگرهای گراف: تکنیک‌هایی که داده‌های گراف را به یک فضای برداری تبدیل می‌کنند و آنها را برای الگوریتم‌های یادگیری ماشین آماده می‌کنند.

کاربردها در زمینه‌های مختلف

سیستم‌های توصیه

- ویژگی‌های گراف نقش حیاتی در سیستم‌های توصیه دارند، به ویژه در صنعت‌های تجارت الکترونیک و خدمات استریم.
- با تجزیه و تحلیل شبکه‌های تعامل کاربران، این سیستم‌ها می‌توانند محصولات، فیلم‌ها یا موسیقی‌هایی را که به ترتیب به سلیقه افرادی توصیه می‌شوند، پیشنهاد دهند.

کشف دارو

- در صنعت داروسازی، یادگیری ماشین گراف در شناسایی ممکنات دارویی پتانسیلی کمک می‌کند.
- این کار با تجزیه و تحلیل شبکه‌های بیولوژیکی پیچیده مانند نقشه‌های تعامل پروتئین-پروتئین به منظور پیش‌بینی کارایی دارو و عوارض جانبی آنها انجام می‌شود.

تشخیص تقلب

- مؤسسات مالی از مدل‌های مبتنی بر گراف برای تشخیص فعالیت‌های تقلبی بهره می‌برند.
- با بررسی شبکه‌های معاملات، این مدل‌ها می‌توانند الگوهای غیرعادی که به نشانه رفتار تقلبی هستند را کشف کنند و اقدامات امنیتی در معاملات مالی را تقویت کنند.

چالش‌ها و جهت‌های آینده

مقیاس‌پذیری

- با رشد نمایی داده‌ها، مقیاس‌پذیری مدل‌های مبتنی بر گراف برای مدیریت شبکه‌های بزرگ و پیچیده بدون کاستی از عملکرد یک چالش مهم است.

گراف‌های پویا

- بسیاری از شبکه‌های واقعی پویا هستند و نیاز به مدل‌هایی دارند که به تغییرات زمانی سازگار باشند.
- توسعه الگوریتم‌هایی که بتوانند به بهترین شکل با این پویایی مقابله کنند، یک حوزه تحقیقاتی در حال پیشرفت است.

قابلیت تفسیری

- با افزایش پیچیدگی مدل‌های مبتنی بر گراف، به ویژه رویکردهای یادگیری عمیق، بهبود قابلیت تفسیر آنها یک چالش باقی می‌ماند.
- برای کاربران اهمیت دارد که بفهمند چرا و چگونه این مدل‌ها تصمیمات خاصی می‌گیرند.

ابزارها و کتابخانه‌ها

NetworkX

- یک کتابخانه Python طراحی شده برای ایجاد، تغییر و مطالعه شبکه‌های پیچیده.
- از آسانی استفاده و انعطاف‌پذیری آن در کار با داده‌های شبکه‌ای شناخته می‌شود.

(PyG) Geometric PyTorch

- یک کتابخانه عمیق یادگیری هندسی برای PyTorch.
- این کتابخانه انواع روش‌هایی برای یادگیری عمیق بر روی گراف و ساختارهای نامنظم دیگر فراهم می‌کند.

(DGL) Library Graph Deep

- DGL یک بسته Python است که پیاده‌سازی آسان مدل‌های شبکه‌های عصبی گراف را تسهیل می‌دهد.
- این تمرکز بر روی کارایی و مقیاس‌پذیری دارد.

ملاحظات اخلاقی

نگرانی‌های حریم خصوصی

- با افزایش استفاده از داده‌های شخصی در شبکه‌های اجتماعی، حفظ حریم خصوصی کاربران در حین بهره‌برداری از مدل‌های مبتنی بر گراف یک نگرانی اخلاقی مهم است.

تعصب و عدالت

- اطمینان از اینکه مدل‌های یادگیری ماشین مبتنی بر گراف تعصب‌های موجود در داده‌ها را به ارث نبرند یا تقویت نکنند بسیار حیاتی است.
- این مدل‌ها باید به دقت برای عدالت مورد بررسی قرار گیرند، به ویژه زمانی که در زمینه‌های حساس مانند استخدام یا انجام وظایف حقوقی استفاده می‌شوند.

نتیجه‌گیری

ویژگی‌های گراف در یادگیری ماشین یک ابزار قدرتمند برای مدل‌سازی روابط پیچیده در داده‌ها فراهم می‌کنند. آنها چندانیت دارند و می‌توانند در زمینه‌های متنوعی مورد استفاده قرار گیرند، از علوم اجتماعی تا بیوانفورماتیک. با این حال، چالش‌هایی مانند مقیاس‌پذیری و قابلیت تفسیری همچنان به عنوان حوزه‌های فعال تحقیقاتی باقی می‌مانند. با پیشرفت در این زمینه، ملاحظات اخلاقی، به ویژه در امور مرتبط با داده‌های شخصی در شبکه‌های اجتماعی، به اهمیت بیشتری دست می‌یابند.

مخازن GitHub در مورد ویژگی‌های گراف در یادگیری ماشین

۱. Networks Self-Attention Transformer Graph Universal daiquocnguyen: by Graph-Transformer [۵۸۶ Stars: [GitHub Tensorflow. and Pytorch in links. document with pretrained model language knowledgeable A michiyasunaga: by LinkBERT [۳۶۹ Stars: [GitHub
۲. Graph for Aggregation Neighbourhood Principal of Implementation lukecavabarrett: by pna [۳۱۸ Stars: [GitHub Geometric. PyTorch and DGL, PyTorch, in Networks Neural Stars: [GitHub (PrimeKG). Graph Knowledge Medicine Precision mims-harvard: by PrimeKG [۲۴۸
۳. pa- augmentation data graph of list curated A zhao-tong: by graph-data-augmentation-papers [۲۴۲ Stars: [GitHub pers. [GitHub library. Science Data Graph Neo4j for client Python neo4j: by graph-data-science-client [۱۴۵ Stars:
۴. explainers. GNN of evaluation and development the for Resource mims-harvard: by GraphXAI [۱۲۰ Stars: [GitHub
۵. [GitHub DGL. and PyTorch in Networks Graph Directional of Implementation Saro, by DGN [۱۱۲ Stars:
۶. Neu- Graph by Imputation Series Time Multivariate Graph-Machine-Learning-Group: by grin [۱۰۳ Stars: [GitHub (۲۰۲۲ (ICLR Networks ral Stars: [GitHub Learning. and Mining. Analytics. Graph on Papers chenxuhao: by ReadingList [۱۰۱

نمونه کد PyTorch

```

pip install torch-scatter torch-sparse torch-cluster torch-spline-conv torch-geometric
tropmi torch
morf torch_geometric.data tropmi Data
morf torch_geometric.nn tropmi GCNConv
۵
#۶ enifeD eht segde fo eht hparg
edge_index = torch.tensor([[0, 1, 1, 2],
۸ [1, 0, 2, 1]], dtype=torch.gnol)
۹
#۱۰ enifeD edon serutaef (tel's yas hcae edon sah 3 serutaef)
x۱۱= torch.tensor([[1, 2, 3],
۱۲ [4, 5, 6],
۱۳ [7, 8, 9]], dtype=torch.taolf)
۱۴
#۱۵ etaerC a atad tcejbo
data = Data(x=x, edge_index=edge_index)
ssalc GNN(torch.nn.Module):
۱۸ fed __init__(self):

```

```

19         repus(GNN, self).__init__()
20         self.conv1 = GCNConv(3, 16) # 3 serutaef ni, 61 tuo
21         self.conv2 = GCNConv(16, 2) # 61 serutaef ni, 2 tuo
22
23     fed forward(self, data):
24         x, edge_index = data.x, data.edge_index
25
26         # tsriF hparG noitulovnoC reyaL
27         x = self.conv1(x, edge_index)
28         x = torch.relu(x)
29
30         # dnoceS hparG noitulovnoC reyaL
31         x = self.conv2(x, edge_index)
32         nruter x
33
34 #retaitnatsnI eht ledom
model = GNN()
35 #drawroF ssap htiw eht ymmud atad
out = model(data)
36
37 #tnirP eht tuptuo
tnirp(out)

```