



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

گزارش سمینار

یادگیری نیمه نظارتی

نگارش

نرگس صالحی
حمیدرضا همتی
ستایش ثانوی
سیدمهدی میرفندرسکی

استاد راهنما

دکتر احسان ناظر فرد

خرداد ۱۴۰۱

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فهرست مطالب

صفحه

عنوان

۱	مقدمه	۱
۲	انواع روش‌های یادگیری	۲
۳	۱-۲ یادگیری تحت نظارت	۳
۳	۲-۲ یادگیری بدون نظارت	۳
۴	۳-۲ یادگیری نیمه نظارتی	۴
۵	۳ روش خودآموزی	۵
۷	۴ نتیجه‌گیری	۷
۸	منابع و مراجع	۸

فهرست اشکال

صفحه

شکل

فصل اول

مقدمه

فصل دوم

انواع روش‌های یادگیری

۱-۲ یادگیری تحت نظارت

از آموزش یک مدل یادگیری ماشین با استفاده از مجموعه داده برچسب‌گذاری شده، یادگیری تحت نظارت یاد می‌شود. برچسب موجود در مجموعه داده از یک یا دو محل زیر تامین می‌شوند:

- تولید برچسب‌ها همراه با تولید داده‌های در مجموعه داده
 - به کارگیری یک متخصص انسانی جهت افزودن برچسب‌ها به داده‌های خام
- از طرفی یادگیری تحت نظارت با وجود دقت بالا چند محدودیت دارد:
- کند بودن (در بعضی موارد نیاز به متخصصان انسانی دارد که تک تک نمونه‌های آموزشی را به صورت دستی برچسب گذاری کنند)
 - پرهزینه بودن (یک مدل باید بر روی حجم زیادی از داده‌های برچسب‌گذاری شده به صورت دستی آموزش داده شود تا پیش بینی‌های دقیق ارائه دهد).

۲-۲ یادگیری بدون نظارت

از سوی دیگر، یادگیری بدون نظارت زمانی است که یک مدل سعی می‌کند الگوهای پنهان، تفاوت‌ها و شباهت‌های داده‌های بدون برچسب را به تنهایی و بدون نظارت انسان استخراج کند. در این روش، داده‌ها بر اساس شباهت‌ها و تفاوت‌ها به خوشه‌ها گروه‌بندی می‌شوند. در حالی که یادگیری بدون نظارت روشی ارزان‌تر برای انجام وظایف آموزشی ارائه می‌دهد، اما معایبی نیز دارد:

- دارای حوزه محدودی از کاربردها (بیشتر برای اهداف خوشه بندی)
- ارائه نتایج با دقت کمتر از یادگیری تحت نظارت

۳-۲ یادگیری نیمه نظارتی

در تعدادی از مسائل دو روش یاد شده به دلیل معایب آن‌ها کارایی نخواهند داشت. در چنین شرایطی یادگیری نیمه نظارتی با ترکیب ویژگی‌های دو روش یاد شده بسیار کمک کننده خواهد بود. به طور خلاصه این روش یک مدل اولیه روی چند نمونه برچسب‌گذاری شده آموزش دیده می‌شود و سپس به طور مکرر بر روی تعداد بیشتری از داده‌های بدون برچسب نتیجه را مشاهده و اعمال می‌کند.

به دلیل ترکیب دو روش یادگیری تحت نظارت و بدون نظارت تفاوت‌هایی با این روش‌ها خواهد داشت:

- این روش برخلاف یادگیری بدون نظارت، برای انواع مشکلات از طبقه بندی و رگرسیون گرفته تا خوشه بندی کار می‌کند.

- این روش برخلاف یادگیری نظارت شده، مقادیر کمی از داده‌های برچسب‌دار و همچنین مقادیر زیادی از داده‌های بدون برچسب استفاده می‌کند که هزینه برچسب‌گذاری دستی و به تبع آن زمان آماده سازی داده‌ها را کاهش می‌دهد.

فصل سوم

روش خودآموزی

یکی از ساده‌ترین روش‌های یادگیری نیمه نظارتی، خودآموزی است. خودآموزی روشی است که در آن می‌توان از هر روش نظارت شده برای طبقه‌بندی یا رگرسیون استفاده کرد و با استفاده از داده‌های برچسب‌گذاری شده و بدون برچسب، آن را به صورت نیمه نظارتی تغییر داد.

روش کار بسیار ساده است. به جای افزودن برچسب‌ها به کل مجموعه داده، تنها بخش کوچکی از داده‌ها را برچسب‌گذاری کرده و از آن برای آموزش یک مدل استفاده می‌کنیم، سپس مدل به بررسی داده‌های بدون برچسب می‌پردازد. اگر روش کار را دقیق‌تر بخواهیم مورد بررسی قرار دهیم، شکل زیر کمک بسیاری به ما خواهد کرد.

۱. ابتدا یک مدل پایه با کمک روش‌های نظارت شده با داده‌های برچسب‌گذاری شده تولید می‌شود.

۲. سپس فرآیندی که تحت عنوان شبه برچسب‌گذاری شناخته می‌شود، اعمال می‌شود. این فرایند شامل پیش‌بینی مدل برای مابقی داده‌های بدون برچسب است. درواقع به ازای هر داده باقی مانده یک شبه برچسب با کمک مدل فعلی تولید می‌شود.

۳. در این مرحله، مطمئن‌ترین پیش‌بینی‌هایی را که مدل تولید کرده در نظر می‌گیریم. با در نظر گرفتن یک سطح اطمینان (مثلاً ۸۰ درصد) اگر هر یک از شبه برچسب‌ها از این سطح اطمینان فراتر رفت، آنها به مجموعه داده برچسب‌گذاری شده اضافه می‌شود. حال یک ورودی داده بزرگتر برای آموزش یک مدل بهبودیافته ایجاد می‌شود.

این فرآیند می‌تواند چندین بار تکرار شود (در اکثر موارد مقدار ۱۰ یک مقدار استاندارد است). در حالی که نمونه‌های موفق از این روش وجود دارد، باید توجه کرد که عملکرد آن ممکن است از یک مجموعه داده برچسب‌گذاری شده به مجموعه دیگر بسیار متفاوت باشد. همچنین موارد زیادی وجود دارد که خودآموزی ممکن است عملکردی بدتر در مقایسه با یادگیری تحت نظارت داشته باشد.

فصل چهارم

نتیجه گیری

منابع و مراجع