

会議報告

The 22nd ACM SIGKDD Conference on Recommender Systems (RecSys 2016)

開催地: Massachusetts Institute of Technology
(ボストン, USA)

開催日程: 2016年9月15日(木) ~ 19日(月)

<http://recsys.acm.org/recsys16/>

1. RecSys 2016

ACM Recommender Systems (略称 RecSys) は、推薦システムを専門とする国際会議である。推薦システムとは、どのものや情報に価値があるかを特定するのを助ける道具である。この会議では、HCI (ヒューマンコンピュータインタラクション)、情報検索、および機械学習などの関連研究を、推薦システムへの応用という観点からまとめている。Chris Anderson の “We are leaving the age of information and entering the age of recommendation.” の言葉に示されるように、推薦システム研究は、既存の分野から独立してきたとの考えに基づき、推薦システム研究が盛んなミネアポリスにて2007年に第1回を開催した。これ以降、2012年の香港を除いて、アメリカとヨーロッパで交互に開催されてきて、第10回の記念となる今回は、USAで5回目開催である。筆者は7回目の参加で、2010年から続けて参加している。開催地は米国のボストンで、会場となったマサチューセッツ工科大学のほか、ハーバード大学など著名な大学が集まった街として知られている。

例年は、5日の開催期間のうち、初日と最後がワークショップ・チュートリアルに割り当てられ、中3日間が本会議であるが、去年に続いて今年も変則的で、最初の2日間はワークショップで、本会議は後半3日間であった。オープニングで発表された参加登録者数は560で、去年の参加者数480や一昨年の500前後より1割ほど増えた。国別やインダストリ・アカデミアの比率などの情報の発表は今年は無かった。日本からの参加者は、筆者が会った限りでは去年とほぼ同じ10名前後だった。去年は日本のアカデミアからの参加は筆者だけという寂しい状況であったが、今年は5名ほど参加していて、学生さんも参加していた。

スポンサーは16社で、去年の10社から大幅に増えた。ダイヤモンドサポーターは検索のGoogle、音楽配信のSpotify、ネット小売りのAlibaba、プラチナサポーターはビデオ配信のNetflix、質問サイトのQuora、レストラン予約サイトのOpenTable、音楽配信のPANDORAと推薦システムが関連するネット企業で占められている。新規のサポーターとしてはMediumとdemandwareのほ

か、おそらく日本からは初めてのシルバーエッグが加わっていた。

2. 招待講演

招待講演は3件あった。最初のClaudia Perlichは、元IBMで、現在のネット広告企業Distilleryに所属している。筆者には、実世界での運用では使えない情報を使ったモデルを学習してしまうleakage問題の形式的定義を試みたKDD 2011のベストペーパーが印象に残っている。今回の講演も、実世界のデータを機械学習で処理するにあたって直面する問題を取り上げた非常に興味深いものであった。分類問題の例をあげ、特徴が一つだけだとクラスの事後確率は0.5を中心に分布し予測は曖昧だが、多数の特徴を用いると0か1を中心に分布し予測は非常に確定的なものになる。しかし、これは訓練データ中の特徴とクラスの対応が確定的で明確な対象がある場合と、これらの対象を確実に分類できるように学習してしまう学習容易性(predictability)バイアスによるものと述べた。このような学習が容易な事例は、分析者が意図した問題とは無関係である場合が多い。講演では、閲覧履歴からWebページの閲覧確率を予測する問題を取り上げ、予測が容易な事例を二つあげていた。一つは、コメントへの書込みなどで目的のサイトへの誘導などを図るポットによるもので、行動が確定的なので容易に予測できる。もう一つは、利用者の指の下に広告を動かして誤ってクリックさせることを誘発するモバイルアプリで、特定のアプリと特定のサイトの訪問の関係は確定的になる。人間の利用者行動をモデル化しようとする場合には、これらの事例を中心に分類する分類器は当然ながら不適切である。以上のように、分類アルゴリズムにデータを単純に与えてしまうのではなく、事例や特徴について十分に考慮することが実データの処理では重要であること説いていた。

二つ目は、GoogleのShashi ThakurによるGoogleNowなどの個人化についての講演であった。内部アルゴリズムについてあまり詳細の講演はなかったが、knowledge graphという知識ベースを活用し、上位概念や同じ概念などを考慮できるようにしていた。

最後はSinan Aralによる社会ネットワークの分析結果をいくつか紹介していた。最初は、評価時に提示する評価値に、利用者の評価が影響されるという結果で、推薦システム分野では著名なCosleyらのSIGCHI 2003の結果と符合するものであった。次はメッセージサービスの利用者が、知人が利用することで自身も利用することになる現象についての分析であった。サービス開始から時間を経るに従って知人の影響は減少するよう見え

るが、交絡因子を考慮すると影響はほぼ定常的であること、新しいものを早期から利用するアーリーアダプタは知人の影響が大きいことなどを紹介していた。また、知人からの影響は婚姻状態に影響され、結婚していると影響されにくいとのことであった。最後は、商品などについての利用者のレビュー文が記名と匿名である場合とを比較し、記名の場合のほうが影響が強かったとの調査結果を照会していた。

3. インダストリアルセッション・チュートリアル・ワークショップ

このRecSysに特徴的なものとして、査読で採録する研究発表とは別に、企業の研究者や技術者を招待しシステムの運用に関する講演をするインダストリアルセッションがある。今年も三つのセッションが一般講演とパラレルに行われた。聴講したセッションのうち、個人的に印象に残ったPinterest社の講演を紹介する。Pinterestは好みの画像を見つけるためのサイトで、画像をピン、画像を自身でまとめたものをボードと呼ぶ。新たな画像を推薦するのに、このボードに基づく類似度を活用して候補を抽出し、閲覧行動とタグ情報などを利用してランキング学習を用いて順位付けしているとのことであった。

チュートリアルは4件で、今年も一般セッションと並列で開催された。ここでは聴講したグループ推薦について紹介する。グループ推薦とは、グループで食事をするときのレストランや、旅行に出かけるときの目的地などを決めるためのものである。グループ自体が嗜好を共有している、目的を共有している、偶然に集まったなどの違いなど系統だった分類を示していた。また、個人の嗜好をどのようにグループの嗜好としてまとめるかについても、和や積などの普通の集約演算に加え、好きなものを推薦されたら、グループの他の人が好きで自分は嫌いなものも推薦されるという公平性など高度なものもあり、個人の嗜好を集約する手法の良いサーベイであった。最初の2日間に10件のワークショップが開催されたが、筆者は本会議のみの参加だったので割愛する。

4. 一般発表・受賞

このRecSysでは一昨年までは一般口頭発表もシングルトラックであったが、採録論文数が増えて昨年からは2パラレルになった。発表はロング、ショート、ポスターに加え、Past, Present and Future (PPF) というポジションペーパーの部門が新設された。ロングとショートは論文のページ数の差で、どちらも優秀なものには口頭発表の機会が与えられる。ポスターは会議録に採録されないもので、意見交換などのために発表する。ロング、ショート、およびPPFの投稿数はそれぞれ、159, 110, 25で、採択数は29, 22, 9であった。採択率は、ロングは昨年の21.4%→18.2%、ショートは26.2%→20%といずれ

も昨年より厳しくなった。国別では米の割合が大きく、伊・仏・中などが多い。日本からも投稿があったと聞くが、採録には至らなかったようだ。この会議は日本からの発表は弱く、第10回になっても口頭発表は1本だけしか採録されていない。

受賞についてまとめておく。ロングとショートのそれぞれの部門で、三つずつの優秀論文があらかじめ選定されており、その中から各部門でのベストペーパーが一つずつ選定される。ロング部門の“Local Item-Item Models for Top-N Recommendation”はアイテム間の類似性に基づく推薦で、利用者の嗜好パターンのグループを考慮することで予測精度を向上させた。ショート部門は“Adaptive, Personalized Diversity for Visual Discovery”で、アイテムの一覧表示画面に選択するものを、アイテムの属性ができるだけ広くカバーされるように選択する。

新設されたPPFについて述べておく。この部門は、研究コミュニティに対し、推薦システム研究と過去と現在に対する認識と、将来のあるべき推薦システムに対するビジョンを示すものである。発表は印象に訴えることを狙ったビジュアルなものから、分析結果などエビデンスを示して主張するものまで幅広かった。全体的に共通するのは利用者中心 (user-centric) ということがキーワードになるようだ。個人的には“HCI for Recommender Systems: the Past, the Present and the Future”が興味深かった。推薦分野の文献のキーワードの割合の変遷を分析した結果を示し、利用者の制御、適応的な計算、感情情報の計算、リスクの見積もりなどの研究が不十分ではないかとの主張をしていた。

個人的に関心のあった一般発表をいくつかあげておく。

- **A Coverage-Based Approach to Recommendation Diversity on Similarity Graph**: 推薦の多様性を集合の被覆問題として捉えて定式化
- **Contrasting Offline and Online Results When Evaluating Recommendation Algorithms**: オンライン評価とオフライン評価では結果が異なるが、新規性と関心が同時に満たされるときに特に顕著
- **Exploring the Value of Personality in Predicting Rating Behaviors**: BigFiveという代表的な性格分類と、映画の好みとの関係を調査
- **Mechanism Design for Personalized Recommender Systems**: ニセ取引で評判システムの評価を向上させないようにするゲーム理論に基づいた枠組み
- **Discovering What You're Known For**: 他者からどのように見られているのかというKnown-forをタグ、位置情報、ソーシャル関係から抽出
- **Intent-Aware Diversification Using a Constrained PLSA**: PLSAモデルのトピックに対する操作により利用者の意図を反映させる

- **ExpLOD: A Framework for Explaining Recommendations based on the Linked Open Data Cloud**: LODの知識の三つ組, 利用者の特性と推薦対象の関係を示すことで説明とする

5. おわりに

基調講演など一部の資料はホームページにて公開されている。会議関連の **Twitter** の **Tweet** は <http://togetter.com/li/1024525> にまとめておいたのでご参考にされたい。

2017年は、イタリアのミラノの北方にあるコモ市で8月27～31日の開催である。会場は18世紀の建築らしく、ホールはオペラ劇場のようで豪華である。イタリアは前述のように、推薦システム研究を続けている大学が多く、研究が盛んであり、今年も多くの興味深い発表が集まることであろう。世界的には注目される会議にもかかわらず日本からの発表や参加が少なく寂しい現状だが、今年こそ日本からも発表や参加が増えればと願う次第である。

[神畠 敏弘 (産業技術総合研究所)]