

会議報告

The 8th ACM Conference on Recommender Systems RecSys 2014

開催地: Crowne Plaza Hotel San Mateo/Foster City
(米国カリフォルニア州フォスターシティ
(シリコン・バレー))

開催日程: 2014年10月6日(月)~10日(金)
<http://recsys.acm.org/recsys14/>

1. RecSys 2014

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、日常的に行われている他人からの推薦に基づく選択を、補助したり、促進したりするものである。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た会議である。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、去年の香港を除いて、アメリカとヨーロッパで交互に開催されてきた。今回の第 8 回は、アメリカのサンフランシスコ湾岸のフォスターシティという場所での開催であった。マウンテンビューやサンタクララなどのシリコンバレーの中心と、サンフランシスコのほぼ中間にあたり、このフォスターシティの周囲にも多くの IT 企業がある。筆者は 5 回目の参加で、2010 年から続けて参加している。スポンサーには、IT 系企業の研究所に加えて、ビデオ配信の Net.ix や音楽配信の PANDORA、そして各ソーシャルネットサービスなど、推薦システムと関連の深い企業も入っている。

2. 会議概要

例年、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。3 件の基調講演、8 件のワークショップ、パネルとコンペティションが 1 件ずつ、そして 4 件のチュートリアルがあった。これらの規模は例年と変わっていない。本会議はシングルセッションであり、すべての発表を聴講することができる。一般の研究発表は、234 件の投稿に対し、採録数は 35 で、採択率は 15% と低かった。例年は投稿数は 150 を超えることはなく、採択率は 20% 前後で推移してきたが、シリコンバレー開催というのも影響したのか、かなり厳しかった。参加者数は 500 人を超えたと発表されていた。これは過去最高

の参加人数であり、この方面の研究が盛んになっていることがうかがえる。

この会議は毎年 RecSys Challenge というデータ分析コンペティションを開催している。今年は、利用者の嗜好予測ではなく、利用者が対象について起こす何らかの行動であるエンゲージメントの予測が対象であった。具体的には、映画の情報サイト IMDb 中の映画について、マイクブログ、Twitter で言及が行われるかどうかを予測する課題であった。

3. 基調講演

基調講演を順に紹介する。最初は、推薦を活用したビデオ配信を行う Netflix 社の Neal Hnut による、推薦を利用した事業についての講演であった。広告や推薦などを見て、関係がないとか、不快に思うような部分があるうちはまだ不完全で、そうしたものがないようなものを最終的に目指す必要があるとのことだった。ビデオ、番組の満足度は視聴時間をフィードバックとしている点などは興味深かった。次は、MapReduce などの開発で著名な Google の Jeff Dean によるものであった。並列計算ではあるが、最近では深層学習に取り組んでいるようである。この講演では、自然言語処理への適用について述べており、深層学習による自然言語データの扱いや、機械翻訳への適用について紹介があった。最後は、Stanford 大学の Hector Garcia-Molina であった。Thoughts on the Future of Recommender Systems というタイトルであったが、スタンフォードで開発された、大学のコースの評価・検索システム CouseRank と、クラウドソーシングを用いた画像検索についての講演であった。

4. チュートリアル・インダストリアルセッション

初日は 4 件のチュートリアルを聴講した。最初の Xavier Amatriain は推薦システム全般について概要であった。新しい話題としては、表示するコンテンツの選択に加え、その配置・見せ方の最適化を試みる研究の紹介があった。2 件目 Huiji Gao による位置ベース SNS での推薦であった。位置ベース情報の空間的・時間的な分布の分析結果についての結果は興味深かった。3 件目は、Paolo Cremonesi による、本と映画など別の分野のアイテムを推薦したりするクロスドメイン推薦の講演であった。タグ推薦などとアルゴリズム的には同じとの認識だったが、同じドメインでも他のドメインの情報を活用したり、ドメイン間で利用者に重複がなかったりする場合など、問題として大きな違いがあることを学ぶことができた。最後は Ido Guy らのソーシャルネットワーク

上の友人推薦などに関する概要であった。

この RecSys に特徴的なものとして、査読により採録された研究発表に加えて、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するインダストリアルセッションがある。6社の招待講演があったので、そのうちいくつかを紹介する。ビジネスを目的とした SNS サービスの LinkedIn 社は、A/B テスト用のインフラである XLNT についての講演であった。クリック率は低下したとき、どの要因が影響していたかを調査したりするために、ランダムサンプリングを行って検定をするのが A/B テストである。これを随時行うためのインフラを昨年あたりから運用を開始し、最新の状況では 1日に 200回以上のテストを、800種以上の尺度に基づいて実行しているとのことであった。ソーシャルネットサービスの Facebook の講演では、嗜好の予測より、コンテンツに対する好意的評価を利用者が示す Like ボタンなどのエンゲージメントの扱いが主であった。Like されるであろうコンテンツを、利用者に知らせることでエンゲージメントの向上を目指したりするとのことであった。推薦ではネット企業が多いが、小売店を対象に個人化サービスを提供する Shopkick の講演では、道路や店舗の配置に基づいて店舗間の移動を分析したりするとのことであった。作付け作物や肥料選択や市場への対応など農業を行ううえで生じる選択に対して情報を提供する Climate 社の講演では、選択に対する推薦について、情報の取得と、その情報の利用のバランスを調整するバンディットの仕組みを利用しているとのことであった。服飾分野での推薦を行う StitchFix では、服飾では重要なデザインに関する情報を扱うために、ヒューマンコンピューテーションを利用して特徴を得ているとのことであった。

5. 一般発表・受賞

一般発表については、上記のように、投稿数が過去最高となり、採択率は非常に低くなった。top-n 推薦、ハイブリッド推薦、行列分解が投稿の上位を占めた分野で、コールドスタート、評価尺度、暗黙フィードバックが採択率の高い分野であった。

ベストペーパーは、Yahoo Labs! の Xing Yi らの Beyond Clicks: DwellTime for Personalization で、ページの閲覧時間を、サーバ側で求めようとするものであった。通常はアクセスログの情報だけのものが多いのだが、コンテンツの長さや内容、閲覧デバイスの種類などの情報も用いて求める。ベストペーパー runner-up は、DePaul 大の Negar Hariri らによる Context Adaptation in Interactive Recommender Systems であった。利用者の嗜好の変化に対して、バンディットを用いて追従できるようにする研究であった。

個人的に関心のあった講演をいくつかあげる。Question Recommendation with Constraints for Massive Open On-line Courses : MOOC で質問する人と解答する人の

マッチングを行う相互推薦で、負荷の量や能力の適性を考慮して、フロー最大化問題として解く。Evaluating Recommender Behavior for New Users : 推薦システムの能力の異なる側面を計測する複数の指標を用いて、アルゴリズムを比較し、アイテム間型は新規利用者に向かないなどの結果を示した。Comparative Recommender System Evaluation: Benchmarking Recommendation Frameworks : 同じアルゴリズムといわれているものでも実装によって、また、ホールドアウトサンプルをどのようにとるかなどの要因によって、評価指標の値が大きく変動することを示し、多様な評価を行うプラットフォーム RiVal を開発した。On Over-Specialization and Concentration Bias of Recommendations: Probabilistic Neighborhood Selection in Collaborative Filtering Systems : 近傍のアイテム・利用者の予測評価値に基づいて利用者の嗜好を予測するが、その近傍の選択を確率的にするという簡潔なアイデアで、特定のアイテムのみが推薦される問題を防ぐ。

6. おわりに

最後に、参加者数などの情報をまとめておく。国内機関からの発表は、今年も本会議ではなく、プロシーディングに収録されないポスターで筆者を含め 2 件、あとはワークショップで 1 件の発表があったのみである。過去 8 回の中でも、本会議のオーラルは 1 件のみ、ショートのパスター発表では 3 件のみという非常に日本の弱さが目立つ会議である。一方で、参加者数は、米 55%、中 5%、日 4% と 3 番目であった。全体では約半分が企業参加者だったが、日本からは 7~8 割は企業参加者であり、アカデミックの関心は低いようである。

本会議の発表のうち、発表者の許諾が得られたものは動画が公開されている*1。講演資料などのリンクなども含んでいる会議関連の Twitter のツイートはまとめサイト*2にまとめておいたので参考にされたい。

2015 年はヨーロッパに移り、オーストリアのウィーンにて 9 月 16~20 日と例年より早い開催である。なお、機械学習・データマイニングの会議 ECMLPKDD はその前週にポルトガルのポルトで開催される。筆者は、ワークショップやポスターなどは毎年発表してはいるが、本会議で発表できるよう、今年も挑戦する予定である。日本の研究機関や企業からも、活発な投稿・参加があればと思うしだいである。

[神島 敏弘 (産業技術総合研究所)]

*1 <https://www.youtube.com/playlist?list=PLaZufLfJumb9A95nS5AmY6G5mqYnwIEZX>

*2 <http://togetter.com/li/699647>