



会議レポート

WSDM 2017 参加報告

「ACM International Conference on Web Search and Data Mining (WSDM)」は、計算機科学の学会組織である「Association for Computing Machinery (ACM)」により2008年から毎年開催されている国際会議であり、「International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)」、「International ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)」、「ACM International Conference on Information and Knowledge Management (CIKM)」、「World Wide Web Conference」とともに情報検索、データマイニング分野に関する難関国際会議として広く認知されている。今回のWSDM 2017は、通算10回目の会議であり、2017年2月6日～10日に英国ケンブリッジギルドホールで開催された。

会期中、初日にStartup day (起業に関するキーノートやパネルディスカッション)、Doctoral consortium および3件のチュートリアルが開催され、次の3日間で本会議が開催された。また最終日にはWSDMで毎年行われるコンペ「WSDM Cup」のワークショップを含め5件のワークショップが開催された。本会議では、採択された80件の論文(すべてフルペーパー)が口頭(long presentationが24件、short presentationが56件)およびポスター(80件すべて)形式で発表された。特にポスターは口頭発表セッション間の休憩時間(午前40分間と午後40分間の2回)に、当日発表されるすべての論文がポスターの形式でも発表され、口頭発表の事前、あるいは、事後にも議論を行うことで論文の理解度を向上させることが可能であった。また、本会議は9つの論文セッション、3つのキーノートセッション、4つのPE (Practice & Experience) キーノートセッションで構成され、すべてのセッションはシングルトラックで行われた。特にWSDMではPE キーノートセッション(主に企業の研究者・役員による招待講演)が、キーノートセッション(主に大学の教授・研究者による招待講演)とは別に設けられており、Google、Microsoft等の企業で行われている研究や最新製品・サービスを紹介する場として用いられている。研究発表以外にも、6日夜にレセプション、

開催年	投稿論文数	採択論文数	採択率
2013	387	73	18.9%
2014	355	64	18.0%
2015	238	39	16.4%
2016	368	67	18.2%
2017	505	80	15.8%

表-1 最近5年間のWSDMの投稿論文数、採択論文数、採択率の推移

8日の夜にバンケットといったイベントも開催され、研究者同士の懇親の場が提供された。

参加者と採択論文の傾向

今回のWSDMの参加者数は379人であり、前回の534人より155人減ったものの、前回に続き歴代2番目の規模であった。また、英国開催ということで通常の国際会議と比べて英国人の割合が多かった。また、発表中には、トランプ(Donald John Trump)大統領の大統領令で米国に再入国できなくなる可能性を恐れた、米国在住の非米国人の欠席、代理発表が目立った。

表-1に最近5年間のWSDMの投稿論文数、採択論文数、採択率をまとめる。今回、採択率は15.8%と最近5年間の平均(約17.5%)より低かったが、投稿論文数は初めて500件を超えたため、採択論文数は近年のWSDMの中で最多となった。また、今回の投稿論文数は2013年の歴代最高値387件を118件超えており、近年、情報検索・データマイニング分野が注目を集めて拡大していることを反映していると考えられる。また投稿論文の半分以上は米国と中国からであり、米国と中国が二強なのは例年通りの傾向となった。日本からの投稿論文は7番目に多く全体投稿数の約3%であり、おそらく15件程度の投稿があったと思われるが、採択論文数は我々の論文も含めて2件であった。奮起が期待される場所である。

図-1に2008年、2011年、2014年、今回のWSDMに発表された論文におけるトピックの上位10件を示す。今回発表された論文全体の特徴としては、ソーシャルネットワーク(social networks)、学習(learning)、推薦システム(recommendation system)、(ニューラルネットワークで用いられる)エンベディング(embedding)の関連論文が増えた点である。特に今回のWSDMではソーシャルネットワーク関連論文がWSDMで伝統的に多かった検索(search)、Web関連論文より多くなっており、SNSの利用増加による研究ニーズの変化を反映していると考えている。また、エンベディング、つまりニューラルネットワークを活用した論文が増えたのも、もう1つの注目点であった。ここ数年、画像・音声・言語処理分野においてニューラルネットワークは輝かしい成果をあげており、ブームと言って良いが、情報検索・データマイニング分野でも同様な動きがあり、今回のWSDMでは10番目に多かったトピックがエンベディングであった。

2008	2011	2014	2017
Search	Web	Search	Social Networks
Web	Search	Web	Search
Ranking	Social	Data	Learning
Mining	Mining	Social Networks	Time
Classification	Query	Modeling	IR
Advertising	Analysis	Learning	Modeling
Models	Temporal	Systems	Text
Document	Online	Click	Data Mining
Graph	Learning	Advertising	Recom. Systems
Analysis	Data	Mining	Embedding

図-1 WSDM に発表された論文におけるトピックの上位 10 件の推移 (Baeza-Yates⁴⁾ より)。

ベストペーパー

今回の WSDM では Garimella ら¹⁾, Joachims ら²⁾, Bendersky ら³⁾ の 3 本の論文がベストペーパーの候補に挙げられ、そのうち Joachims らがベストペーパーに、Garimella らがベストスチューデントペーパーに選定された。WSDM の創設者の 1 人である Baeza-Yates 氏が、今回の招待講演⁴⁾ で「情報検索とデータマイニング分野における今後の課題」の 1 つとして挙げたのは、多様な要因で生じる「Web データ中の偏り (bias in the Web)」であった。選定された論文は Web データの偏り問題を解決するための手法を提案した点が高く評価されたと思われる。

Joachims らは文書検索結果を再ランキングする際に用いるユーザのクリックデータには、「検索結果においてランキングの高い文書のうち、検索に対して適切な文書はその多くがクリックされるが、より低いランキングの適切な文書は多くの場合クリックされない」という偏りがあり、そうした偏りのあるクリックデータをそのまま再ランキングの学習に用いても大きな性能向上は期待できないという問題に着目した。Joachims 氏は元の検索結果におけるランキングにクリックされる傾向を反映した上で、再ランキングを学習させる position-based propensity model を提案し、クリックデータにおける偏りの問題解決を試みた。また、提案モデルは Joachims 氏が長年開発してきた svm-light 上で実装され、「Propensity SVMrank」というツールとして一般に公開されている^{☆1}。

Garimella らは twitter 等、SNS 上での議論における

echo chamber 現象 (SNS 上では同じ観点を持つ人同士がよく繋がっていて、この繋がりの中では同じ観点の発言や議論がよく共有されるが、違う観点からの発言や議論は少ない、あるいは拒否される傾向がある) によって複数ある異論のうち 1 つのみがユーザに共有される偏り問題が生じると指摘した。Garimella 氏は SNS 上の発言を極性付きノードに対応させ、リツイート等のユーザ間のインタラクションをグラフに表現した。また、各々の極性におけるグラフ上の highest degree node (つまり、リツイートが多い発言) を結び付け、ユーザに提示することで偏り問題を解決しようとした。

WSDM Cup 2017

WSDM では毎年 WSDM Cup というコンペが行われ、成績上位 3 位までのチームに表彰と賞金が授与される。今回は「vandalism detection task」(Wiki データに誤った情報の掲載等コンテンツにダメージを与える修正が行われた場合、それを自動検出するタスク) と「triple scoring task」(「Jonny Depp:: 職業 :: 俳優」のような意味的關係 triple の正しさを判定するタスク) の 2 つのタスクでコンペが行われた。日本からはベンチャー企業の Ousia と NAIST の合同チームが triple scoring task に参加し、2 位の好成績をあげて賞金 \$750 を授与された。

次回の WSDM

今回の WSDM は、世界最大のマリーナといわれる米国のロサンゼルスのマリーナ・デル・レイ (カリフォルニア州) で 2018 年 2 月 5 日～9 日の 5 日間開催される予定である。

参考文献

- 1) Garimella, K. : Gianmarco De Francisci Morales, Aristides Gionis, Michael Mathioudakis. Reducing Controversy by Connecting Opposing Views. *In Proceedings of WSDM 2017*, pp.81-90 (2017).
- 2) Joachims, T. : Adith Swaminathan, Tobias Schnabel. Unbiased Learning-to-Rank with Biased Feedback. *In Proceedings of WSDM 2017*, pp.781-790 (2017).
- 3) Bendersky, M., Wang, X., Metzler, D., Najork, M : Learning from User Interactions in Personal Search via Attribute Parameterization. *In Proceedings of WSDM 2017*, pp.791-800 (2017).
- 4) Baeza-Yates, R. : Ten Years of Wisdom. *In Proceedings of WSDM 2017*, pp.1-2 (2017).

(呉 鍾勲/国立研究開発法人 情報通信研究機構)

☆1 http://www.joachims.org/svm_light/svm_proprank.html