Explaining Recommendations: Fidelity versus Interpretability

Derek Bridge
Insight Centre for Data Analytics
University College Cork, Ireland
Overview

• Recommender Systems
• Explaining Recommendations
• Case Studies
• Concluding Remarks
RECOMMENDER SYSTEMS
What is a Recommender System?

- Software that helps users **discover**
  - new music and other media
  - cultural artefacts such as works of art and architecture
  - products and services
  - travel experiences
  - ...

- Recommendations must typically be
  - relevant to the user (**personalized**) and the context-of-use (**contextualized**)
  - diverse
  - serendipitous
  - ...

Photo by Nickolai Kashirin (CC by 2.0)
A Scenario

A hungry academic ....

...receives a recommendation for a place-to-eat but

...not within walking distance

...a fusion-style cuisine with which the academic is unfamiliar.

...Her confidence in the recommendation might be improved by an explanation...
A hungry academic ....

...receives a recommendation for a place-to-eat but

...not within walking distance

...a fusion-style cuisine with which the academic is unfamiliar.

...Her confidence in the recommendation might be improved by an explanation...
Types of Recommender System

• Content-based
• Collaborative
  – User-based nearest-neighbours
  – Item-based nearest-neighbours
  – Matrix factorization
Types of Recommender System

- Content-based
- Collaborative
  - User-based nearest-neighbours
  - Item-based nearest-neighbours
  - Matrix factorization

Training set → Build Model → User & Context-of-use
Types of Recommender System

- Content-based
- Collaborative
  - User-based nearest-neighbours
  - Item-based nearest-neighbours
  - Matrix factorization
Content-Based

Crime, drama

Adventure, drama, fantasy

Western

Action, sci-fi

Comedy, drama, romance
## User-Based Nearest-Neighbours

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
<td></td>
<td>4</td>
<td>4</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>5</td>
<td>2</td>
<td></td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>1</td>
<td></td>
<td></td>
<td>5</td>
<td></td>
</tr>
</tbody>
</table>
# User-Based Nearest-Neighbours

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>User</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>User</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>User</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

**User-user similarity**

![user1.png](image1).jpg  ![user2.png](image2).jpg  ![user3.png](image3).jpg  ![user4.png](image4).jpg  ![user5.png](image5).jpg  ![user6.png](image6).jpg  ![user7.png](image7).jpg
<table>
<thead>
<tr>
<th>Item-Based Nearest-Neighbours</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="1" alt="Image" /></td>
<td><img src="2" alt="Image" /></td>
<td><img src="3" alt="Image" /></td>
<td><img src="4" alt="Image" /></td>
<td><img src="5" alt="Image" /></td>
<td><img src="6" alt="Image" /></td>
</tr>
<tr>
<td><img src="1" alt="Image" /></td>
<td><img src="2" alt="Image" /></td>
<td><img src="3" alt="Image" /></td>
<td><img src="4" alt="Image" /></td>
<td><img src="5" alt="Image" /></td>
<td><img src="6" alt="Image" /></td>
</tr>
<tr>
<td><img src="1" alt="Image" /></td>
<td><img src="2" alt="Image" /></td>
<td><img src="3" alt="Image" /></td>
<td><img src="4" alt="Image" /></td>
<td><img src="5" alt="Image" /></td>
<td><img src="6" alt="Image" /></td>
</tr>
<tr>
<td><img src="1" alt="Image" /></td>
<td><img src="2" alt="Image" /></td>
<td><img src="3" alt="Image" /></td>
<td><img src="4" alt="Image" /></td>
<td><img src="5" alt="Image" /></td>
<td><img src="6" alt="Image" /></td>
</tr>
<tr>
<td><img src="1" alt="Image" /></td>
<td><img src="2" alt="Image" /></td>
<td><img src="3" alt="Image" /></td>
<td><img src="4" alt="Image" /></td>
<td><img src="5" alt="Image" /></td>
<td><img src="6" alt="Image" /></td>
</tr>
</tbody>
</table>
# Item-Based Nearest-Neighbours

<table>
<thead>
<tr>
<th></th>
<th>Item-item similarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
</tr>
</tbody>
</table>

- **Item 1** is similar to Item 5.
- **Item 2** is similar to Item 4.
- **Item 3** is similar to Item 5.
- **Item 4** is similar to Item 3.
- **Item 5** is similar to Item 4.
Matrix Factorization

$n$ movies

$m$ users
Matrix Factorization

$n$ movies

$m$ users

$f$ latent factors

$m$ users

$n$ movies

$f$ latent factors
Matrix Factorization

\[ n \text{ movies} \]

\[ m \text{ users} \]

\[ f \text{ latent factors} \]

\[ m \text{ users} \]

\[ n \text{ movies} \]

\[ f \text{ latent factors} \]
Ever More Complex Models

- Hybrids and Ensembles
- Multi-Objective Systems
- Deep Models
- Latent Feature Spaces
Interpretable Models

- Intelligible global descriptions of systems
  - E.g. decision trees
  - E.g. linear models (esp. sparse linear models)

- Challenges
  - preserving accuracy
  - intelligibility, e.g. when there are many features or highly-engineered features
  - protecting Intellectual Property

- Interpretable deep models
  - learn to associate semantic feature with nodes in hidden layers
DARPA’s XAI Initiative

B.1 Explainable Models

New Approach

Create a suite of machine learning techniques that produce more explainable models, while maintaining a high level of learning performance.

Learning Techniques (today)

- Neural Nets
- Deep Learning
- Graphical Models
- Bayesian Belief Nets
- Ensemble Methods
- Random Forests
- Decision Trees
- SRL
- CRFs
- HDNs
- MLNs
- AOGs
- SVMs
- Markov Models

Deep Explanation

Modified deep learning techniques to learn explainable features.

Interpretable Models

Techniques to learn more structured, interpretable, causal models.

Model Induction

Techniques to infer an explainable model from any model as a black box.

https://www.darpa.mil/program/explainable-artificial-intelligence
EXPLAINING RECOMMENDATIONS
Explanations are Relational
Explanations are Relational

• Recommendation only
  – “You might like *Never Let Me Go*”
Explanations are Relational

• Recommendation only
  – “You might like *Never Let Me Go*”

• Recommendation plus description
  – “You might like *Never Let Me Go*, a 2010 dystopian drama based on the 2005 novel of the same name...”
Explanations are Relational

• Recommendation only
  – “You might like Never Let Me Go”

• Recommendation plus description
  – “You might like Never Let Me Go, a 2010 dystopian drama based on the 2005 novel of the same name...”

• Recommendation plus explanation
  – “You liked Atonement, so you might also like Never Let Me Go”
Intermediaries in Explanations

User

Items

Users

Features

Recommendation

[Vig et al., 2009]
Intermediaries in Explanations

User \(\xrightarrow{\text{likes}}\) Items

Users

Features

Recommendation

[Vig et al., 2009]
Intermediaries in Explanations

User

likes

Items

are similar to

Recommendation

Users

Features

[Vig et al., 2009]
Intermediaries in Explanations

User likes Items are similar to Recommendation

User is similar to Users is similar to Features

[Vig et al., 2009]
Intermediaries in Explanations

Vig et al., 2009

User

likes

is similar to

Items

are similar to

Users

who like

Recommendation

Features
Intermediaries in Explanations

User \(\rightarrow\) is similar to \(\rightarrow\) Users \(\rightarrow\) who like \(\rightarrow\) Recommendation

Items \(\rightarrow\) are similar to

Features \(\rightarrow\) likes

User likes \(\rightarrow\) is similar to \(\rightarrow\) Users

[Vig et al., 2009]
Intermediaries in Explanations

User \(\rightarrow\) is similar to \(\rightarrow\) Users \(\rightarrow\) who like \(\rightarrow\) Recommendations

Users \(\rightarrow\) are similar to \(\rightarrow\) Items

Features \(\rightarrow\) likes \(\rightarrow\) User

[User] \(\rightarrow\) likes \(\rightarrow\) Items

[Vig et al., 2009]
Explanation Dimensions

- Interpretable
- Ethical
- Sound and Complete (Fidelity)
- Actionable
- Cheap-to-compute
Fidelity

- **Soundness**: How truthful each element in an explanation is with respect to the underlying system.
- **Completeness**: The extent to which an explanation describes all of the underlying system.

\[ \text{Fidelity} = \text{Soundness} + \text{Completeness} \]

[Kulesza et al., 2013]
Fidelity

Soundness
How truthful each element in an explanation is with respect to the underlying system

Completeness
The extent to which an explanation describes all of the underlying system

Fidelity

increasing trust, fewer requests for clarification, better understanding

[Kulesza et al., 2013]
White-Box Explanations

Training set → Build Model → User & Context-of-use
White-Box Explanations

Training set → Build Model → Recommendation + “trace” data

User & Context-of-use
White-Box Explanations

Training set → Build Model → Explanation Generation

User & Context-of-use

Recommendation + “trace” data

Recommendation + Explanation
White-Box Explanations

Sound explanations

Training set → Build Model → User & Context-of-use

Recommendation + “trace” data

Explanation Generation

Recommendation + Explanation
Black-Box Explanations

Training set → Build Model → User & Context-of-use
Black-Box Explanations

Training set ➔ Build Model ➔ Explanation Generation ➔ Recommendation ➔ Recommendation + Explanation

User & Context-of-use
Black-Box Explanations

1. Training set
2. Build Model
3. User & Context-of-use
4. Explanation Generation
5. Recommendation
6. Recommendation + Explanation
Black-Box Explanations

- User & Context-of-use
  - Training set
  - Other data
  - Build Model
  - Explanation Generation
  - Recommendation
  - Recommendation + Explanation
Black-Box Explanations

Training set → Build Model → Explanation Generation

Other data

User & Context-of-use

Queries

Recommendation

Recommendation + Explanation
Black-Box Explanations

Model-agnostic, probably not sound

Training set

Build Model

Other data

Explanation Generation

Queries

Recommendation

Recommendation + Explanation

User & Context-of-use
Why Explain?

- Scrutability
- Trust
- Persuasion
- Decision support
CASE STUDIES
CASE STUDY A

White-Box Explanations of User-Based Nearest-Neighbours Recommendations
Explaining User-Based Nearest Neighbours Recommendations

- Difficulties
  - Often 50+ neighbours
  - Userids are meaningless: strangers!
  - Profiles are large and private
Explaining User-Based Nearest Neighbours Recommendations

• Difficulties
  – Often 50+ neighbours
  – Userids are meaningless: strangers!
  – Profiles are large and private

  – Good at persuading [Herlocker et al, 2000]
  – Less good for trust and decision-support [Bilgic & Mooney, 2005]
Explaining User-Based Nearest Neighbours Recommendations

Training set → Build Model → Recommendation + Neighbours → Explanation Generation → Recommendation + Explanation
CASE STUDY B

Item-Based Explanations for User-Based Nearest-Neighbours Recommendations
Item-Based Explanations

- Good for trust and decision-support [Bilgic & Mooney, 2005]
- Familiar, e.g. Amazon:
  - “Customers who bought Atonement also bought Never Let Me Go”
Item-Based Explanations for User-Based Recommendations

User is similar to Users who like Recommendation

User

is similar to

Users

who like

Recommendation

NEVER LET ME GO
Item-Based Explanations for User-Based Recommendations

- **Explanation partners**
  - The user’s neighbours

- **Candidate items**
  - The movies the user has in common with her partners

- **Association rules**
  - Rules that link candidates to the recommended item

[Bridge & Dunleavy, 2014; Kaminskas, Durão & Bridge, 2017]
Item-Based Explanations for User-Based Recommendations

1. Training set
2. Build Model
3. User & Context-of-use
4. Recommendation + neighbours
5. Explanation Generation
6. Recommendation + Explanation
Item-Based Explanations for User-Based Recommendations

Training set → Build Model → Explanation Generation

- Mine association rules
- Recommendation + neighbours
- Recommendation + Explanation

User & Context-of-use
Using Queries in Black-Box Explanations

(but these two case studies are not recommender systems!)
Queries in Black-Box Explanations

Training set → Build Model → Explanation Generation

User & Context-of-use

Queries

Recommendation

Recommendation + Explanation
Are You Safe To Drive?

[Doyle et al, 2004; Bridge & Cummins, 2006]
Are You Safe To Drive?

[Doyle et al, 2004; Bridge & Cummins, 2006]
Are You Safe To Drive?

[Doyle et al, 2004; Bridge & Cummins, 2006]
Are You Safe To Drive?

[Doyle et al, 2004; Bridge & Cummins, 2006]
Are You Safe To Drive?

[Doyle et al, 2004; Bridge & Cummins, 2006]
LIME: Explanations for any Classifier

Training set → Build Model → Build Local Model

\[ \arg \min_{g \in G} \mathcal{L}(f, g, \Pi_x) + \Omega(g) \]

Classification + Explanation

[Ribeiro et al, 2016]
Opinionated Recommendation
Opinionated Recommendation

- Reviews as source of features & sentiment
- Recommendations that balance similarity & sentiment
- Novel explanation generation
- Re-ranking of recommendations

“Great Location, Average Hotel”

We stayed here as part of a family ski trip in March 2015. The hotel was part of a Topflight package and we upgraded to a large room with a balcony (2 adults and one 4 year old).

The hotel is in a very good location. Westendorf is small but the hotel is very close to the nursery slopes (about 50m) and ski shops. So getting to the snow was effortless every day.

As for the hotel it is very average. Our room as large but very basic. It was comfortable but would benefit hugely from a makeover.

The food was also very average. Typical Austrian affair which may or may not be your cup of tea. I guess this is what one signs up for in Westendorf. It would have been very easy to improve the breakfast however, which was very limited.

Free wifi was pretty decent but a bit flakey at times but better than most I have experienced in ski resorts.

[Muhammad et al., 2016]
Opinionated Recommendation

Muhammad et al., 2016
Opinionated Recommendation

Clontarf Castle Hotel

1,985 Reviews | #20 of 174 Hotels in Dublin | Certificate of Excellence

Reasons for you to choose this hotel:
- Bar/Lounge (better than 60% of alternatives)
- Free Parking (better than 90% of alternatives)
- Restaurant (better than 70% of alternatives)

Reasons for you to avoid this hotel:
- Airport Transportation (worse than 90% of alternatives)
- Leisure Centre (worse than 75% of alternatives)

This explanation has been generated based on things that matter to you. Click here to see additional features.

[Muhammad et al., 2016]
CASE STUDY F

Recommendation-By-Explanation
Recommendation-By-Explanation

[Rana & Bridge, 2017]
Recommendation-By-Explanation

The Notebook
- star-crossed-lovers
- secret-love
- broken-engagement
- volunteer
- u.s.-army
- romantic-rivalry
- self-discovery

Candidate

[Rana & Bridge, 2017]
Recommendation-By-Explanation

User’s past preferences

The Illusionist
• fiancé-fiancee relationship
• shooting
• secret-love
• broken-engagement
• star-cross-lovers

The Notebook
• star-crossed-lovers
• secret-love
• broken-engagement
• volunteer
• u.s.-army
• romantic-rivalry
• self-discovery

[User’s past preferences]

[The Illusionist]

[The Notebook]

[Reference: Rana & Bridge, 2017]
Recommendation-By-Explanation

User’s past preferences

The Illusionist
- fiancé-fiancée relationship
- shooting
- secret-love
- broken-engagement
- star-cross-lovers

The Notebook
- star-crossed-lovers
- secret-love
- broken-engagement
- volunteer
- u.s.-army
- romantic-rivalry
- self-discovery

Candidate

[Rana & Bridge, 2017]
Recommendation-By-Explanation

Pearl Harbour
- fiancé-fiancee relationship
- shooting
- secret-mission
- volunteer
- parachute
- ...

The Illusionist
- fiancé-fiancee relationship
- shooting
- secret-love
- broken-engagement
- star-cross-lovers

The Notebook
- star-crossed-lovers
- secret-love
- broken-engagement
- volunteer
- u.s.-army
- romantic-rivalry
- self-discovery

User’s past preferences

Candidate

[Rana & Bridge, 2017]
Recommendation-By-Explanation

Pearl Harbour
- fiancé-fiancée relationship
- shooting
- secret-mission
- volunteer
- parachute

The Illusionist
- fiancé-fiancée relationship
- shooting
- secret-love
- broken-engagement
- star-cross-lovers

The Notebook
- star-crossed-lovers
- secret-love
- broken-engagement
- volunteer
- u.s.-army
- romantic-rivalry
- self-discovery

User’s past preferences

Candidate

[Rana & Bridge, 2017]
Recommendation-By-Explanation

Pearl Harbour
- fiancé-fiancée relationship
- shooting
- secret-mission
- volunteer
- parachute

The Illusionist
- fiancé-fiancée relationship
- shooting
- secret-love
- broken-engagement
- star-cross-lovers

The Notebook
- star-crossed-lovers
- secret-love
- broken-engagement
- volunteer
- u.s.-army
- romantic-rivalry
- self-discovery

User’s past preferences

[Rana & Bridge, 2017]
Recommendation-By-Explanation

Big Fish
- romantic-rivalry
- carnival
- secret-mission
- parachute
- ...

Pearl Harbour
- fiancé-fiancée relationship
- shooting
- secret-mission
- volunteer
- parachute
- ...

The Illusionist
- fiancé-fiancée relationship
- shooting
- secret-love
- broken-engagement
- star-cross-lovers

The Notebook
- star-crossed-lovers
- secret-love
- broken-engagement
- volunteer
- u.s.-army
- romantic-rivalry
- self-discovery

User’s past preferences

Candidate

[Rana & Bridge, 2017]
Recommendation-By-Explanation

Big Fish
- romantic-rivalry
- carnival
- secret-mission
- parachute
- ...

Pearl Harbour
- fiancé-fiancée relationship
- shooting
- secret-mission
- volunteer
- parachute
- ...

The Illusionist
- fiancé-fiancée relationship
- shooting
- secret-love
- broken-engagement
- star-cross-lovers

The Notebook
- star-crossed-lovers
- secret-love
- broken-engagement
- volunteer
- u.s.-army
- romantic-rivalry
- self-discovery

User’s past preferences

Candidate

[Rana & Bridge, 2017]
Recommendation-By-Explanation

- Big Fish
  - romantic-rivalry
  - carnival
  - secret-mission
  - parachute
  - ...

- Pearl Harbour
  - fiancé-fiancée relationship
  - shooting
  - secret-mission
  - volunteer
  - parachute
  - ...

- The Illusionist
  - fiancé-fiancée relationship
  - shooting
  - secret-love
  - broken-engagement
  - star-cross-lovers

- The Notebook
  - star-crossed-lovers
  - secret-love
  - broken-engagement
  - volunteer
  - u.s.-army
  - romantic-rivalry
  - self-discovery

User’s past preferences

Candidate

[Rana & Bridge, 2017]
Recommendation-By-Explanation

Precision

Diversity

Surprise
Comparison

Classical approaches
- Generate & rank recommendations
- Generate explanations

Opinionated Recommendation
- Generate recommendations
- Generate explanations
- Re-rank recommendations

Recommendation -By-Explanation
- Generate reasons to recommend (explanations)
- Recommend those with the best reasons
CONCLUDING REMARKS
The Future

• Drive explanation design & evaluation by explanation goals
• Bring explanations into the heart of recommenders

• Design explanations that go beyond relevance to the user
  – context-aware
  – diversity
  – serendipity
The Future

• Design ways to present and visualise recommendations and explanations

• Design explanations for systems that are more conversational
  – why are you recommending this?
  – what aren’t you recommending that?
  – why this ahead of that?
  – what would you recommend if things were different? (counterfactuals)
Acknowledgements

Lisa Cummins
Dónaíl Doyle
Fred Durãº
Arpit Rana

Padraig Cunningham
Kevin Dunleavy
Marius Kaminskas

and all other members of the Recommender Systems Group, Insight Centre for Data Analytics
References

- Bridge, Derek and Dunleavy, Kevin: If you liked Herlocker et al.’s explanations paper, then you might like this paper too, Proceedings of the Workshop on Interfaces and Human Decision Making for Recommender Systems (Worshop Programme of the Eighth ACM Conference on Recommender Systems), pp.22-27, 2014
- Doyle, Doyle
- Vig, Jesse and Sen, Shilad and Riedl, John: Tagsplanations: Explaining Recommendations Using Tags, Proceedings of the 14th International Conference on Intelligent User Interfaces, pp.47—56, 2009
Obrigado!