
Lea Frermann
August 24, 2017

joint work with Gyuri Szarvas
A Short History of Books

- Ancient Egypt: 3000 BC
- Ancient Greece: 2000 BC
- Roman Empire: 0 AD
- Middle Ages: 500 AD

Library of Alexandria: ‘concurrent copying’
A Short History of Books

Book Production in Europe

Data from https://ourworldindata.org/books/
Book Production in Europe

Data from https://ourworldindata.org/books/
Book Production in Europe

- 6th century AD: 120 books per year
- 7th century AD: 120 books per year
- 8th century AD: 120 books per year
- 9th century AD: 120 books per year
- 10th century AD: 120 books per year
- 11th century AD: 120 books per year
- 12th century AD: 120 books per year
- 13th century AD: 120 books per year
- 14th century AD: 120 books per year
- 15th century AD: 20 million books per year
- 16th century AD: 20 million books per year
- 17th century AD: 20 million books per year
- 18th century AD: 20 million books per year

Data from https://ourworldindata.org/books/
What should I read next?

New U.S. Fiction Titles, 1940–2010

What should I read next?

New U.S. Fiction Titles, 1940–2010
What should I read next?

New U.S. Fiction Titles, 1940–2010
What should I read next?

'I want a book about space pirates'

'I want a book about a secret-baby romance'

New U.S. Fiction Titles, 1940–2010

Year

What should I read next?

New U.S. Fiction Titles, 1940–2010

Popular Tags on List

Displaying Tag count 31 - 60 of 1000 in totals

- christian (569)
- adventure (540)
- paranormal-romance (509)
- titles (466)
- 2014 (436)
- 2015 (410)
- urban-fantasy (408)
- science (401)
- 2016 (393)
- humor (369)
- middle-grade (559)
- adult (535)
- books (503)
- poetry (456)
- manga (425)
- setting (410)
- biography (405)
- comics (398)
- contemporary-romance (384)
- art (368)
- vampires (373)
- novel (365)

→ previous 1 2 3 4 5 6 7 8 9 ... 33 34 next →
What should I read next?

New U.S. Fiction Titles, 1940–2010

Book Fiction Moods
- Action-packed
- Military

Science Fiction & Fantasy Genres
- Thriller
- Non-Romantic
- Mystery
- Horror
- Humor
- Romantic

Book Mystery Characters
- Police Officers
- Amateur Sleuths
- Private Investigators
- Government Agents

Year
- 1940
- 1950
- 1960
- 1970
- 1980
- 1990
- 2000
- 2010
## Book tags the Amazon Catalog

<table>
<thead>
<tr>
<th>Title</th>
<th>Price</th>
<th>Format</th>
<th>Rating</th>
<th>Views</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Secret Baby</em> (601)</td>
<td>$2.99 to buy</td>
<td>Auto-delivered wireless</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Colby (BBW Western Bear Shifter Romance) (Rodeo Bears Book 3)</em></td>
<td>$0.00 kindleunlimited</td>
<td>Kindle Edition</td>
<td>4.5/5</td>
<td>145</td>
</tr>
<tr>
<td><em>Fury: A Secret Baby Romance</em></td>
<td>$0.00 kindleunlimited</td>
<td>Kindle Edition</td>
<td>4.5/5</td>
<td>43</td>
</tr>
<tr>
<td><em>Football Baby: A Secret Baby Romance</em></td>
<td>$0.00</td>
<td>Free with Audible trial</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Motivation and Goals

Goal: Flexible, fine-grained book recommendation

- induce **multi-view** representations of books
- views encode **relevant** and **distinct** information
- use representations to group books into locally coherent groups (**micro-genres**)
Motivation and Goals

But, how do we know our clusters are meaningful?

- books that share gold tags are similar
- a scalable and principled evaluation framework
Prior Work
Prior work – Unsupervised learning of plot representations

- Typical event chains (McIntyre and Lapata, 2010)
- Typical characters (Bamman et al., 2014; Elsner, 2012)
- Social network analysis (Elson et al., 2010)
- Character relations and their development (Iyyer et al., 2016)
Prior work – Unsupervised learning of plot representations

- **Typical event chains** (McIntyre and Lapata, 2010)
- **Typical characters** (Bamman et al., 2014; Elsner, 2012)
- **Social network analysis** (Elson et al., 2010)
- **Character relations and their development** (Iyyer et al., 2016)

**Relationship Modeling Networks (RMN)**

We extend RMNs to induce multi-view representations and introduce a new loss function.
Prior work – Evaluation

- manual analysis
- crowd-sourced interpretability judgments
- small manual gold-standards of character types
- distinguish natural vs. scrambled stories

Elsner (2012); Bamman et al. (2013, 2014); Iyyer et al. (2016),...

→ *intrinsic* measures
→ utility for downstream tasks?
→ artificial data
Prior work – Evaluation

- manual analysis
- crowd-sourced interpretability judgments
- small manual gold-standards of character types
- distinguish natural vs. scrambled stories

Elsner (2012); Bamman et al. (2013, 2014); Iyyer et al. (2016), ...

→ intrinsic measures
→ utility for downstream tasks?
→ artificial data

We propose a general, scalable and principled framework utilizing Amazon catalog book tags.
Prior work – Language-based recommendation

Prevalent recommendation paradigm: social filtering

- bags-of-words from book information databases
- semantic frames from book plot summaries
- topic modelling for scientific article recommendation

Mooney and Roy (1999); Clercq et al. (2014); Wang and Blei (2011),...
Prior work – Language-based recommendation

Prevalent recommendation paradigm: social filtering

- bags-of-words from book information data bases
- semantic frames from book plot summaries
- topic modelling for scientific article recommendation

Mooney and Roy (1999); Clercq et al. (2014); Wang and Blei (2011),...

We use richer information from the full book text, and learn structured multi-view representations
MVPlot – A Model of Multi-View Plot Representations
The Idea

Character Properties

view v1 characteristics of individual characters
The Idea

- enmity
- companionship
- love
- friendship
- competition
- hate
- time
- young
- shy
- threatened
- brave
- hero
- smart
- shy
- lonely
- lucky
- rich
- bold
- injured
- stranded
- hungry
- dead

Character Properties

Character Relations

view v1 characteristics of individual characters
view v2 relations between character pairs
Properties and Relations: word clusters (≈ topics)
Objective: triggers views to capture distinct information
v1 corpus

\[ s_t \] 20-word text spans mentioning only character c1

\{ s_t \} chronologically ordered sequence of spans \{ s_1, s_2, \ldots, s_T \}
MVPlot: Input

Text spans \{s_t\}

s_1[... c1 ....... ..... ... ]
s_2[... ..... c1... ..... ...]
s_3[c1 ..... ...... ... ]
s_4[... ...... ... c1 ...]
s_t[... ........ ........ c1 ...]

v_1 corpus

v_2 corpus

v_1 \ s_t \ 20\text{-word text spans mentioning only character c1}

v_2 \ s_t \ 20\text{-word text spans mentioning only characters c1 and c2}

\{s_t\} chronologically ordered sequence of spans \{s_1, s_2, \ldots, s_T\}
For many books and many characters / character pairs

\[ s_t \]

20-word text spans mentioning only character c1

\[ s_t \]

20-word text spans mentioning only characters c1 and c2

\[ \{ s_t \} \]

chronologically ordered sequence of spans \( \{ s_1, s_2, \ldots, s_T \} \)
• map words in spans to pre-trained GloVe embedding
• average word embeddings to obtain span embedding
1. A set of global **property descriptors** v1, e.g.,

<table>
<thead>
<tr>
<th>despair</th>
<th>alone</th>
<th>desperate</th>
<th>sad</th>
</tr>
</thead>
<tbody>
<tr>
<td>intellect</td>
<td>smart</td>
<td>clever</td>
<td>knowledge</td>
</tr>
<tr>
<td>look</td>
<td>pretty</td>
<td>small</td>
<td>slim</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

2. A set of global **relation descriptors** v2, e.g.,

<table>
<thead>
<tr>
<th>love</th>
<th>dearest</th>
<th>loving</th>
<th>attracted</th>
<th>like</th>
</tr>
</thead>
<tbody>
<tr>
<td>crime</td>
<td>murder</td>
<td>chase</td>
<td>escape</td>
<td>prosecute</td>
</tr>
<tr>
<td>hate</td>
<td>hatred</td>
<td>dislike</td>
<td>fear</td>
<td>anger</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
3. For each **character**: a **property trajectory**
   → a sequence of probability distributions over \( v1 \) descriptors

4. For each **character pair**: a **relationship trajectory**
   → a sequence of probability distributions over \( v2 \) descriptors
MVPlot: Intuition

Model architecture based on Iyyer et al (2016)
Model architecture based on Iyyer et al (2016)
MVPlot: Intuition

Model architecture based on Iyyer et al (2016)
MVPlot: Intuition

Properties, Relations change smoothly over time
hidden rep. at time $t$ depends on hidden rep. at time $t - 1$

Model architecture based on Iyyer et al (2016)
MVPlot: Architecture

**multi-view loss**

- **single view v1**
  - reconst span embedding
  - descr weights $t - 1$
  - descr weights $t$
  - $d_{v1}^{t-1}$
  - $d_{v1}^t$
  - $R_{v1}$
  - $r_{v1}^t$
  - $v_{1}$ dictionary

- **pair view v2**
  - reconst span embedding
  - descr weights $t$
  - descr weights $t - 1$
  - $d_{v2}^t$
  - $d_{v2}^{t-1}$
  - $R_{v2}$
  - $r_{v2}^t$
  - $v_{2}$ dictionary

- **input**
  - span embedding
  - $e^t$

- **hidden input representation**
  - $h^t$
The Multi-View Loss
1) Within each view,
   → parameters that efficiently represent and **accurately reconstruct** the input

2) Across views,
   → parameters that encode relevant and **distinct aspects**
Term 1: View-specific Hinge Loss with Negative Sampling

1) Within each view,

→ parameters that **accurately reconstruct** the input
1) Within each view,

→ parameters that **accurately reconstruct** the input

\( v_1 \) parameters reconstruct **true** \( v_1 \) input **well**

\( v_1 \) parameters reconstruct **random** \( v_1 \) inputs **poorly**

\[
J(\theta) = \sum_t \beta \sum_{n \in N} \max(0, 1 - r^t_{v_1} e^t_{v_1} + r^t_{v_1} e^n_{v_1}) + (1 - \beta) \max(0, 1 - e^t_{v_1} r^t_{v_1} + e^t_{v_1} r^t_{v_2}) + \lambda \| R_{v_1} R^T_{v_1} - I \|
\]

**Loss Term 1 based on Iyyer et al (2016)**
1) Within each view,

→ parameters that **accurately reconstruct** the input

\( v_2 \) parameters reconstruct **true** \( v_2 \) input **well**

\( v_2 \) parameters reconstruct **random** \( v_2 \) inputs **poorly**

\[ J(\theta) = \sum_t \beta \sum_{n \in N} \max(0, 1 - r_{v_1}^t e_{v_1}^t + r_{v_1}^t e_{v_1}^n) + (1 - \beta) \max(0, 1 - e_{v_1}^t r_{v_1}^t + e_{v_1}^t r_{v_2}^t) + \lambda \| R_{v_1} R_{v_1}^T - I \| \]

**Loss Term 1 based on Iyyer et al (2016)**
Term 2: Cross-View Max-Margin Loss

2) Across views,

→ parameters that encode relevant and **distinct aspects**
Term 2: Cross-View Max-Margin Loss

2) Across views,

→ parameters that encode relevant and distinct aspects

\( v_1 \)-parameters reconstruct true \( v_1 \) input well

\( v_2 \)-parameters reconstruct true \( v_1 \) input poorly

\[
J(\theta) = \sum_t \beta \sum_{n \in N} \max(0, 1 - r_{t,v_1}^{e_t} e_{v_1} + r_{t,v_1}^{n_i} e_{v_1}) + (1 - \beta) \max(0, 1 - e_{v_1}^{t} r_{t,v_1}^{e_t} + e_{v_1}^{t} r_{t,v_2}^{e_t}) + \lambda \| R_{v_1} R_{v_1}^T - I \|
\]

maximize margin between true input and negative samples

maximize margin between true view’s and other view’s reconstruction

Regularizer
Term 2: Cross-View Max-Margin Loss

2) Across views,

→ parameters that encode relevant and distinct aspects
  \( v_2 \)-parameters reconstruct true \( v_2 \) input well
  \( v_1 \)-parameters reconstruct true \( v_2 \) input poorly

\[
J(\theta) = \sum_t \beta \sum_{n \in N} \max(0, 1 - r_{v_1}^t e_{v_1}^t + r_{v_1}^t e_{v_1}^n) + (1 - \beta) \max(0, 1 - e_{v_1}^t r_{v_1}^t + e_{v_1}^t r_{v_2}^t) + \lambda \| R_{v_1} R_{v_1}^T - I \|)
\]

maximize margin between true input and negative samples

maximize margin between true view’s and other view’s reconstruction

Regularizer
Evaluation
Evaluation Overview

Data

<table>
<thead>
<tr>
<th></th>
<th>Amazon corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td># books</td>
<td>10,000</td>
</tr>
<tr>
<td># v1 sequences</td>
<td>91,511</td>
</tr>
<tr>
<td># v2 sequences</td>
<td>70,156</td>
</tr>
</tbody>
</table>
## Data

<table>
<thead>
<tr>
<th></th>
<th>Amazon corpus</th>
<th>Gutenberg corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td># books</td>
<td>10,000</td>
<td>3,500</td>
</tr>
<tr>
<td># v1 sequences</td>
<td>91,511</td>
<td>45,182</td>
</tr>
<tr>
<td># v2 sequences</td>
<td>70,156</td>
<td>60,493</td>
</tr>
</tbody>
</table>
## Evaluation Overview

### Data

<table>
<thead>
<tr>
<th></th>
<th>Amazon corpus</th>
<th>Gutenberg corpus</th>
</tr>
</thead>
<tbody>
<tr>
<td># books</td>
<td>10,000</td>
<td>3,500</td>
</tr>
<tr>
<td># v1 sequences</td>
<td>91,511</td>
<td>45,182</td>
</tr>
<tr>
<td># v2 sequences</td>
<td>70,156</td>
<td>60,493</td>
</tr>
</tbody>
</table>

- Preprocessing with BooksNLP pipeline (Bamman et al., 2014)
- keep sequences $s$ iff $5 \leq |s| \leq 250$
Evaluation of quality of induced descriptors
Experiment 1 – Evaluation of induced Descriptors

Data: Gutenberg corpus
Descriptors represented as 10 closest words in GloVe space

<table>
<thead>
<tr>
<th>Property Descriptors (v1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>laugh scream laughing yell joke cringe disgrace embarrassment hate cursing</td>
</tr>
<tr>
<td>snug fleece warm comfortable wet blanket flannel cozy comfort roomy</td>
</tr>
<tr>
<td>excellency mademoiselle monsieur majesty duchess empress madame countess madam dowager</td>
</tr>
</tbody>
</table>

love loving lovely dear sweetest dearest thank darling congratulation hello associate assistant senior chairman executive leadership vice director liaison vice-president
Experiment 1 – Evaluation of induced Descriptors

Data: Gutenberg corpus
Descriptors represented as 10 closest words in GloVe space

<table>
<thead>
<tr>
<th>Property Descriptors (v1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>laugh scream laughing yell joke cringe disgrace embarrassment</td>
</tr>
<tr>
<td>hate cursing</td>
</tr>
<tr>
<td>snug fleece warm comfortable wet blanket flannel cozy comfort</td>
</tr>
<tr>
<td>roomy</td>
</tr>
<tr>
<td>excellency monsieur majesty duchess empress madame countess madam dowager</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Relation Descriptors (v2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>love loving lovely dear sweetest dearest thank darling congratulation hello</td>
</tr>
<tr>
<td>associate assistant senior chairman executive leadership vice director liaison vice-president</td>
</tr>
</tbody>
</table>
Experiment 1 – Evaluation of induced Descriptors

Do $v^2$ relation descriptors indeed capture character relations?

- MVPlot $v^2$ descriptors vs RMN descriptors (Iyyer et al., 2016)
- crowd-sourced plausibility judgments (Chaturvedi et al., 2017)
Do \( v_2 \) relation descriptors indeed capture character relations?

- MVPlot \( v_2 \) descriptors vs RMN descriptors (Iyyer et al., 2016)
- crowd-sourced plausibility judgments (Chaturvedi et al., 2017)

<table>
<thead>
<tr>
<th>love</th>
<th>loving</th>
<th>lovely</th>
<th>dear</th>
<th>sweetest</th>
<th>dearest</th>
<th>thank</th>
<th>darling</th>
<th>congratulation</th>
<th>hello</th>
</tr>
</thead>
</table>

Do the words describe a relation, event or interaction between people?

- yes
- no
Experiment 1 – Evaluation of induced Descriptors

Gutenberg corpus

% of annotators who marked descriptor relevant

% descriptors marked relevant

0 20 40 60 80

80 60 40 20 0

RMN Gutenberg
MVPlot Gutenberg

RMN Amazon
MVPlot Amazon

22 / 30
Experiment 1 – Evaluation of induced Descriptors

Gutenberg corpus and Amazon corpus

Gutenberg corpus and Amazon corpus

% of annotators who marked descriptor relevant

% descriptors marked relevant

% of annotators who marked descriptor relevant

RMN Gutenberg

MVPlot Gutenberg

RMN Amazon

MVPlot Amazon

22 / 30
Large-scale evaluation of local neighborhoods in model space
Experiment 2: Large-scale evaluation of book neighborhoods

Are novels with similar induced representations indeed similar?

- Quantify the quality of induced micro-clusters
- Utilize expert-tags from the Amazon catalog: do similar books share tags?
Experiment 2: Large-scale evaluation of book neighborhoods

Are novels with similar induced representations indeed similar?

- Quantify the quality of induced micro-clusters
- Utilize expert-tags from the Amazon catalog: do similar books share tags?

<table>
<thead>
<tr>
<th>Genre</th>
<th>Refinement: Character Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thriller</td>
<td>British Detectives; FBI Agents; Female Protagonists</td>
</tr>
<tr>
<td>Romance</td>
<td>Cowboys; Criminals &amp; Outlaws; Doctors; Spies; Wealthy</td>
</tr>
<tr>
<td>SciFi</td>
<td>AIs; Aliens; Clones; Mutants; Psychics; Robots &amp; Androids</td>
</tr>
</tbody>
</table>
Experiment 2: Large-scale evaluation of book neighborhoods

Are novels with similar induced representations indeed similar?

- Quantify the quality of induced micro-clusters
- Utilize expert-tags from the Amazon catalog: do similar books share tags?

<table>
<thead>
<tr>
<th>Genre</th>
<th>Refinement: Mood and Theme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Thriller</td>
<td>Action-packed; dark; fun; racy &amp; risque</td>
</tr>
<tr>
<td>Romance</td>
<td>Second-chances; secret-baby; international; workplace</td>
</tr>
<tr>
<td>SciFi</td>
<td>horror; humor; mystery; non-romantic</td>
</tr>
</tbody>
</table>
Experiment 2: Large-scale evaluation of book neighborhoods

We select 50 tags

- from popular genres (thriller, romance, scifi)
- frequent refinements (character, category, mood/theme)
- Not tuned at all.
Experiment 2: Large-scale evaluation of book neighborhoods

Each book $b$ is represented through two views

$v_1$ For each character
   property representation $c$ (averaged trajectory)

$v_2$ For each character pair
   relation representation $p$ (averaged trajectory)
Each book $b$ is represented through two views

$v_1$ For each character
    property representation $c$ (averaged trajectory)

$v_2$ For each character pair
    relation representation $p$ (averaged trajectory)

We can cluster books based on different views and ask

- Which books have similar character types ($v_1$)?
- Which books have similar character pair relations ($v_2$)?
- Which books are similar overall ($v_1, v_2$)?
Experiment 2: Large-scale evaluation of book neighborhoods

Each book $b$ is represented through two views

$v_1$ For each character
  property representation $c$ (averaged trajectory)

$v_2$ For each character pair
  relation representation $p$ (averaged trajectory)

View-specific book similarities, given books $b_1, b_2$

$$sim(v_1) = \sum_{c \in b_1} \sum_{c' \in b_2} \min(dist_{euclid}(c, c'))$$

$$sim(v_2) = \sum_{p \in b_1} \sum_{p' \in b_2} \min(dist_{euclid}(p, p'))$$

$$sim(v_1, v_2) = \frac{1}{2}(sim(v_1) + sim(v_2))$$
Experiment 2: Large-scale evaluation of book neighborhoods

Setup

- Reference set: 500 most popular books in Amazon dataset (by # reviews)
- For each, compute the 10 nearest neighbors
- Do nearest neighbors share a gold tag with the reference? (precision@10)
### Results

<table>
<thead>
<tr>
<th>Model</th>
<th>View</th>
<th>P@10</th>
<th>MAP</th>
</tr>
</thead>
<tbody>
<tr>
<td>MVPlot</td>
<td>$v_1$</td>
<td>0.529</td>
<td>0.401</td>
</tr>
<tr>
<td></td>
<td>$v_2$</td>
<td>0.496</td>
<td>0.367</td>
</tr>
<tr>
<td></td>
<td>$v_1, v_2$</td>
<td><strong>0.546</strong></td>
<td><strong>0.421</strong></td>
</tr>
<tr>
<td>RMN</td>
<td>$v_2$</td>
<td>0.479</td>
<td>0.347</td>
</tr>
<tr>
<td></td>
<td>$v_1$</td>
<td>0.516</td>
<td>0.392</td>
</tr>
<tr>
<td>cosine</td>
<td>$v_2$</td>
<td>0.468</td>
<td>0.339</td>
</tr>
<tr>
<td></td>
<td>$v_1, v_2$</td>
<td>0.512</td>
<td>0.390</td>
</tr>
</tbody>
</table>

Significance with $p < 0.05$ (†) or $p < 0.01$ (‡) (paired t-test).
Conclusions
Conclusions

MVPlot – Deep multi-view book representations

- learns sets of **property** and **relation** descriptors
- efficient on both large-scale (Amazon) and smaller (Gutenberg) data
- multi-view model performs best
  → views capture *distinct* information
Conclusions

Multi-view Loss function

- novel multi-view loss function
- triggers views to capture **relevant** and **distinct** information
- generalizes to arbitrary # views and different input data
Conclusions

Evaluation of plot representations

- task-based extrinsic evaluation
- scalable and principled evaluation framework
Thank you!

www.frermann.de
References


