Recommender system with serendipity

Clémence Réda
supervised by Claire Vernade

Abstract

A lacking feature of current recommender systems is that they usually do not allow discovery of new elements, that is, an element that might be interesting, but which is different of what the user is used to. Moreover, a clear definition of serendipity is missing. The goal of this project was (1) to formalize the problem of recommendation allowing serendipities, that is, surprisingly good discoveries; (2) to design a method which solves this problem in an online setting; (3) to evaluate its relevance with respect to the random strategy, usual collaborative filtering recommender system, and methods using Rotting Bandits.

1 Introduction And Review of the State-of-the-Art

TODO

2 Formalization of the Problem of Recommendation With Serendipity

Serendipities are objects that are unexpectedly enjoyable. A more formal definition would be, given a serendipity threshold s, a unweighted, undirected similarity graph $G(V, E)$ on the objects, and a given user u to which recommendations would be made, which rating values belong to $\mathcal{R} \subset \mathbb{N}$, serendipities are objects that are located in unexplored (with respect to the considered user) object regions (i.e. strongly connected components of the graph), at distance (in terms of length of the shortest path from explored regions to the considered unexplored region) lesser than t, such that these objects can increase in expectation the diameter of the explored regions of the similarity graph by the user across time. Formally, at time t, the set of serendipity objects for user u is as follows:

$$S_u^t = \underset{v \in V - \text{Support}(f_u^{(t)}), d(v, \text{Support}(f_u^{(t)})) \leq s}{\arg\max} \mathbb{E}(\sum v' \in V[f_u^{t+1}(v") - f_t - u(v")] r_u(v")) \quad (1)$$

where $f_u^{(t)} : V \to \{0, 1\}$ is a function that indicates exploration of an object by user u (at time t), $r_u : V \to \mathcal{R}$, a function of rating by user u of objects, d is the distance measure used to build G and, for any subset C of objects, $d(., C) = \min_{c \in C} d(., c)$, and s the so-called "serendipity threshold".

Intuitively, they can be seen as objects, which are not in already explored regions at time t, and which give the greatest increase in the diameter of explored regions of the graph, weighted by the known ratings given by the user u.

where E is the expectation: exploration and rating of a node are drawn from a random distribution: when an object is recommended, we "make" the user explore the recommended node, and this modifies the exploration/rating distributions (if the user has enjoyed the recommended object, they will most likely be willing to explore the neighbours of this object). f-u would likely be submodular

32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montréal, Canada.

(a node can be explored only once, thus at a given point in time, the number of explored elements will increase less and less strongly).

3 Method

The idea would be to use the OIMP (Online Influence Maximization with Persistence) method developped in [Lagrée et al., 2017]. The concept of restricting the set of candidates to a number K of objects (with their supports) might suit the idea to restrict recommendations to non-explored suitable regions:

For a given user u, a fixed parameter K (number of candidates) For each round 1 up to a finite horizon T Compute the set of relevant candidates c-1, ..., c-k with their support A-1, ..., A-k (*) Compute a score of relevance for each candidate Select the (1) candidate with the highest score Observe reward r (rating of the recommended object by user u) Write the associated regret: real best candidate reward (among the k selected candidates) - r Improve the algorithm: update statistics accordingly for score computation Return cumulative regret up to horizon T for comparison with other methods In order to check the serendipity part, we can also keep track of the exploration of each object region: if the recommendation has succeeded in increasing the exploration of the user, the exploration of each region should increase: a mean to quantify this is to compute the volume of the parallelotope which relies on explored (by user u) objects o-1, o-2, ..., o-n as follows: V = (o-1, ..., o-n) and Vol(o-1, ..., o-n) = sqrt(det(V*V-T)) . This volume should increase more for recommendation with serendipity than for regular recommendation, because recommended objects are supposed to be less correlated (similar). [1]

The goal of this algorithm is then to infer the real edge weights (which corresponds, for an edge (i, j), to the probability of user u liking object i (j) knowing its rating of object j (i)) in an online setting.

(*) This is a difficult problem. Lagrée et al. suggest Max-degree or Cover as general strategies to find plausible candidates. One idea is to implement serendipity at this point of the bandit algorithm: select 1 "centroid" (either by K-means, or by Max-degree) in each unexplored object region at distance at least s from the explored regions.

Another issue is the missing data in most of the datasets (MovieLens included). We can restrict movies to objects having a minimal number of ratings, and, for each movie, the reward is a noisy observation of the empirical mean of the ratings in the dataset (using multinomial or Gaussian distributions): "semi-simulated data".

We implicitly assume that the number of "centroids" will be far lesser than the number of nodes in the object graph. When computing the regret, we also assume that the best "arm" we could use is among the K selected candidates, thus that we do not question the method with which we select the candidates, which is an hypothesis made in the [Lagrée et al., 2017] article.

4 Experiments

Comparison with Random strategy, if time permits, with Rotting Bandits and Linear UCB. Horizon T = 1,000. Comparison of the expected cumulative regret and evolution of volume up to horizon T.

5 Discussion

NeurIPS requires electronic submissions. The electronic submission site is

`https://cmt.research.microsoft.com/NeurIPS2018/`

Please read the instructions below carefully and follow them faithfully.

5.1 Style

Papers to be submitted to NeurIPS 2018 must be prepared according to the instructions presented here. Papers may only be up to eight pages long, including figures. Additional pages *containing only acknowledgments and/or cited references* are allowed. Papers that exceed eight pages of content

(ignoring references) will not be reviewed, or in any other way considered for presentation at the conference.

The margins in 2018 are the same as since 2007, which allow for $\sim 15\%$ more words in the paper compared to earlier years.

Authors are required to use the NeurIPS LaTeX style files obtainable at the NeurIPS website as indicated below. Please make sure you use the current files and not previous versions. Tweaking the style files may be grounds for rejection.

5.2 Retrieval of style files

The style files for NeurIPS and other conference information are available on the World Wide Web at

`http://www.neurips.cc/`

The file `neurips_2018.pdf` contains these instructions and illustrates the various formatting requirements your NeurIPS paper must satisfy.

The only supported style file for NeurIPS 2018 is `neurips_2018.sty`, rewritten for LaTeX 2_ε. **Previous style files for LaTeX 2.09, Microsoft Word, and RTF are no longer supported!**

The LaTeX style file contains three optional arguments: `final`, which creates a camera-ready copy, `preprint`, which creates a preprint for submission to, e.g., arXiv, and `nonatbib`, which will not load the `natbib` package for you in case of package clash.

New preprint option for 2018 If you wish to post a preprint of your work online, e.g., on arXiv, using the NeurIPS style, please use the `preprint` option. This will create a nonanonymized version of your work with the text "Preprint. Work in progress." in the footer. This version may be distributed as you see fit. Please **do not** use the `final` option, which should **only** be used for papers accepted to NeurIPS.

At submission time, please omit the `final` and `preprint` options. This will anonymize your submission and add line numbers to aid review. Please do *not* refer to these line numbers in your paper as they will be removed during generation of camera-ready copies.

The file `neurips_2018.tex` may be used as a "shell" for writing your paper. All you have to do is replace the author, title, abstract, and text of the paper with your own.

The formatting instructions contained in these style files are summarized in Sections 6, 7, and 8 below.

6 General formatting instructions

The text must be confined within a rectangle 5.5 inches (33 picas) wide and 9 inches (54 picas) long. The left margin is 1.5 inch (9 picas). Use 10 point type with a vertical spacing (leading) of 11 points. Times New Roman is the preferred typeface throughout, and will be selected for you by default. Paragraphs are separated by ½ line space (5.5 points), with no indentation.

The paper title should be 17 point, initial caps/lower case, bold, centered between two horizontal rules. The top rule should be 4 points thick and the bottom rule should be 1 point thick. Allow ¼ inch space above and below the title to rules. All pages should start at 1 inch (6 picas) from the top of the page.

For the final version, authors' names are set in boldface, and each name is centered above the corresponding address. The lead author's name is to be listed first (left-most), and the co-authors' names (if different address) are set to follow. If there is only one co-author, list both author and co-author side by side.

Please pay special attention to the instructions in Section 8 regarding figures, tables, acknowledgments, and references.

7 Headings: first level

All headings should be lower case (except for first word and proper nouns), flush left, and bold.

First-level headings should be in 12-point type.

7.1 Headings: second level

Second-level headings should be in 10-point type.

7.1.1 Headings: third level

Third-level headings should be in 10-point type.

Paragraphs There is also a `\paragraph` command available, which sets the heading in bold, flush left, and inline with the text, with the heading followed by 1 em of space.

8 Citations, figures, tables, references

These instructions apply to everyone.

8.1 Citations within the text

The `natbib` package will be loaded for you by default. Citations may be author/year or numeric, as long as you maintain internal consistency. As to the format of the references themselves, any style is acceptable as long as it is used consistently.

The documentation for `natbib` may be found at

`http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf`

Of note is the command `\citet`, which produces citations appropriate for use in inline text. For example,

```
\citet{hasselmo} investigated\dots
```

produces

Hasselmo, et al. (1995) investigated…

If you wish to load the `natbib` package with options, you may add the following before loading the `neurips_2018` package:

```
\PassOptionsToPackage{options}{natbib}
```

If `natbib` clashes with another package you load, you can add the optional argument `nonatbib` when loading the style file:

```
\usepackage[nonatbib]{neurips_2018}
```

As submission is double blind, refer to your own published work in the third person. That is, use "In the previous work of Jones et al. [4]," not "In our previous work [4]." If you cite your other papers that are not widely available (e.g., a journal paper under review), use anonymous author names in the citation, e.g., an author of the form "A. Anonymous."

8.2 Footnotes

Footnotes should be used sparingly. If you do require a footnote, indicate footnotes with a number[1] in the text. Place the footnotes at the bottom of the page on which they appear. Precede the footnote with a horizontal rule of 2 inches (12 picas).

[1] Sample of the first footnote.

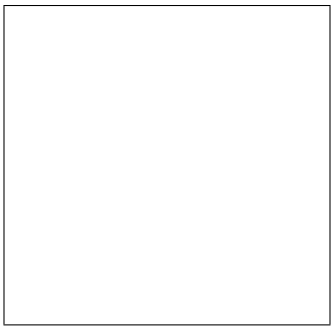

Figure 1: Sample figure caption.

Table 1: Sample table title

	Part	
Name	Description	Size (μm)
Dendrite	Input terminal	$\sim$100
Axon	Output terminal	$\sim$10
Soma	Cell body	up to 10^6

Note that footnotes are properly typeset *after* punctuation marks.[2]

8.3 Figures

All artwork must be neat, clean, and legible. Lines should be dark enough for purposes of reproduction. The figure number and caption always appear after the figure. Place one line space before the figure caption and one line space after the figure. The figure caption should be lower case (except for first word and proper nouns); figures are numbered consecutively.

You may use color figures. However, it is best for the figure captions and the paper body to be legible if the paper is printed in either black/white or in color.

8.4 Tables

All tables must be centered, neat, clean and legible. The table number and title always appear before the table. See Table 1.

Place one line space before the table title, one line space after the table title, and one line space after the table. The table title must be lower case (except for first word and proper nouns); tables are numbered consecutively.

Note that publication-quality tables *do not contain vertical rules.* We strongly suggest the use of the `booktabs` package, which allows for typesetting high-quality, professional tables:

`https://www.ctan.org/pkg/booktabs`

This package was used to typeset Table 1.

9 Final instructions

Do not change any aspects of the formatting parameters in the style files. In particular, do not modify the width or length of the rectangle the text should fit into, and do not change font sizes (except perhaps in the **References** section; see below). Please note that pages should be numbered.

[2] As in this example.

10 Preparing PDF files

Please prepare submission files with paper size "US Letter," and not, for example, "A4."

Fonts were the main cause of problems in the past years. Your PDF file must only contain Type 1 or Embedded TrueType fonts. Here are a few instructions to achieve this.

- You should directly generate PDF files using `pdflatex`.
- You can check which fonts a PDF files uses. In Acrobat Reader, select the menu Files>Document Properties>Fonts and select Show All Fonts. You can also use the program `pdffonts` which comes with `xpdf` and is available out-of-the-box on most Linux machines.
- The IEEE has recommendations for generating PDF files whose fonts are also acceptable for NeurIPS. Please see `http://www.emfield.org/icuwb2010/downloads/IEEE-PDF-SpecV32.pdf`
- `xfig` "patterned" shapes are implemented with bitmap fonts. Use "solid" shapes instead.
- The `\bbold` package almost always uses bitmap fonts. You should use the equivalent AMS Fonts:

  ```
  \usepackage{amsfonts}
  ```

  followed by, e.g., `\mathbb{R}`, `\mathbb{N}`, or `\mathbb{C}` for $\mathbb{R}$, $\mathbb{N}$ or $\mathbb{C}$. You can also use the following workaround for reals, natural and complex:

  ```
  \newcommand{\RR}{I\!\!R} %real numbers
  \newcommand{\Nat}{I\!\!N} %natural numbers
  \newcommand{\CC}{I\!\!\!\!C} %complex numbers
  ```

  Note that `amsfonts` is automatically loaded by the `amssymb` package.

If your file contains type 3 fonts or non embedded TrueType fonts, we will ask you to fix it.

10.1 Margins in LaTeX

Most of the margin problems come from figures positioned by hand using `\special` or other commands. We suggest using the command `\includegraphics` from the `graphicx` package. Always specify the figure width as a multiple of the line width as in the example below:

```
\usepackage[pdftex]{graphicx} ...
\includegraphics[width=0.8\linewidth]{myfile.pdf}
```

See Section 4.4 in the graphics bundle documentation (`http://mirrors.ctan.org/macros/latex/required/graphics/grfguide.pdf`)

A number of width problems arise when LaTeX cannot properly hyphenate a line. Please give LaTeX hyphenation hints using the `\-` command when necessary.