Literature Review

June 8, 2016

Contents

Introduction

This paper outlines the assumptions, strengths and weaknesses of the main types of recommender systems. Challenges in real-world applications are highlighted, including handling sparse, high-dimensional data such as user likes. State of the art implementations for systems are identified, including recent advancements in learning representations.[1] Finally, the future of recommender systems is discussed, showing its relevance to the team's design project.

[1] Yoshua Bengio, Aaron Courville, and Pierre Vincent. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1798–1828, 2013

Collaborative Filtering

Collaborative filtering systems leverage natural overlaps in user preferences to recommend new content to users.[2] The main idea is that if a group of users engage mostly with the same set of content, their tastes are likely similar. Engagement data can be extracted from user searchers, ratings and customized lists. It can also be indirectly measured through views and time spent.

[2] Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46: 109–132, 2013

Strengths

Collaborative filtering systems do not require additional data[3] about the content being recommended. This is advantageous when such data is unreliable or difficult to understand, such as suggesting paintings based on visual features.[2]

[3] Examples include creation date, content type, user tags, user comments and text associated with an item.

Weaknesses

The strength of a recommendation depends on how many users have engaged with the content. Items that are already popular are much easier to recommend. Undiscovered items cannot be suggested directly, including newly added content, which is the core of many services. This is known as the **cold start** problem.[2]

Implementation

There are two approaches to collaborative filtering: modelling users and modelling items. Starting with user engagement, the system can find similar users. On the other hand, similar items can be found. An item to item approach may need to be taken in large datasets, especially when the number of users greatly exceeds the number of items.[2]

At the core of both techniques is a **similarity metric** and thus the development of a vector space. In the user to user approach, users collectively vote for item recommendations, with closer users having a stronger voice. Similarly, in the item to item case, items close to the items that a user already likes can be recommended. Common metrics used are Pearson similarity and cosine similarity.[2]

Allowing every single user or item to vote is impractical for large datasets. Instead, **Neighbourhood models**, are used to enable only the nearest vectors, under some decision rule, to vote on a recommendation.[2] These models are heuristic-based and are largely dependent on finding good representations[4] of data, a topic discussed in greater detail in the dimensionality reduction section. Increasing the size of a neighbourhood significantly increases the complexity of a model. Nearest neighbours approaches have poor scalability for extremely large datasets.[2]

[4] One can think of representations as a machine learned vector space that items and users live in.

Neighbourhood models, which use a weighted voting structure, fall under the classification of **memory-based** techniques for collaborative filtering. In contrast, **model-based** approaches attempt to predict the probabilities of outcomes. The arguably simplest probability model is the **Naive Bayes** classifier, which assumes that the probability of voting is conditionally independent of classes. Such classes are often unobserved, but could include user demographics. Several clustering algorithms, such as k-means, can be used to find latent[5] classes.[6] Many other models exist, including **Bayesian networks**, **artificial neural networks**, and **Markov decision** processes.

[5] Latent, in the context of machine learning, refers to unobserved representations of data that are "learned" by algorithms.

[6] John S Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998

State of the Art

One of the most common approaches for a model-based system is performing **Matrix Factorization**, a process that naturally estimates features by minimizing the reconstruction error of an engagement matrix. Optimization is usually done through iterative, stochastic reduction of a loss function with L2 **regularization** to prevent overfitting.[7]

[7] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008

State of the art systems incorporate latent variables like **Temporal Dynamics** into matrix factorization. These approaches model changes in user behaviour or product trends over time. Authors of this technique argue that most rating values are due to effects independent of user-item interactions.[7] This is particularly important for emotion-based recommender systems, due to the varying nature of mood.

A classic example of latent factors is that some users have a tendency to like more items than other users. These effects, which are not user-item interactions, are modeled via **baseline predictors**. Thus interactions are modeled as a product of factorized user and item matrices plus baseline effects. Further, the matrix factorization problem is improved by adding latent variables and corresponding loss terms to account for implicit relationships in engagement data. This can capture, for example, the fact that users who rate an item were searching for it to begin with.[7]

These advancements to matrix factorization techniques enabled the researchers to outperform benchmarks in the Netflix Dataset.[7] Other authors found that using **Restricted Boltzmann Machines**, a simple two-layer undirected graphical model, also beat benchmarks.[8]

Very recently, the **Word2Vec** (W2V) model, discussed later in the context of text mining, was used as an efficient algorithm for embedding items or users into a lower-dimensional vector space.[9] The approach outperformed Singular Value Decomposition, one of the most common dimensionality reduction techniques used for collaborative filtering.[10]

[8] Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798. ACM, 2007

[9] Oren Barkan and Noam Koenigstein. Item2vec: Neural item embedding for collaborative filtering. *arXiv preprint arXiv:1603.04259*, 2016

[10] Daniel Billsus and Michael J Pazzani. Learning collaborative information filters. In *ICML*, volume 98, pages 46–54, 1998

Challenges

Depending on how polarized user engagement is (do most users like the same content?), the output of models can become unsophisticated. Another risk is over-fitting based on past history, making it hard for users to change their preferences.

Furthermore, since the actual content of items is not understood, there can be great variance in a recommendation set from a model. This makes presenting the information to users difficult. Moreover, it can risk exposing inappropriate content, especially in the case of journals.

Content-based Filtering

Content-based recommender systems find items that are similar to ones a user liked in the past.[2] Various signals[11] like term frequency[12] are useful for finding similar pieces of text.[13,14] Sophisticated signals, such as models of the sentiment and topics in text, can also be used as features.

In such a system, the profile of a user can be constructed as a weighted combination of item features,[14] where the weight is the importance of each feature to the user.[15] This combination can be found by finding a weighted sum of items rated by the user or more sophisticated techniques that compute probabilities.[16]

[11] A signal is a variable in a dataset or a modeled variable such as the topic of a journal.

[12] More accurately, it is standard to use **TF-IDF**, which normalizes term frequencies by their document frequencies.

[13] Zan Huang, Wingyan Chung, Thian-Huat Ong, and Hsinchun Chen. A graph-based recommender system for digital library. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 65–73. ACM, 2002

[14] Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007

[15] An example would be a user who reads posts about a particular topic, where the topic is a feature.

[16] These two approaches should seem very similar to the ones in collaborative filters.

Strengths

In a content-based system, user feedback can be used directly to improve recommendations. Based on user responses, the system can learn to rank and personalize specific features in items.[14] For example, one approach would be to adjust the weight of features making up a user profile to allow for fast customization. In a collaborative filtering approach, it is less straight forward to personalize an individual user's model.

Weakness

Content-based systems face challenges in understanding complex mediums like images and text. Secondly, the systems tend to overspecialize, since their objective is to find similar content.[2] Interestingly, one can optimize for novelty or contrast in items to help users find unique content.

Implementation

Common models utilize **Logistic Regression**, **Support Vector Machines** (SVM), **Decision Trees** (and their **Random Forest** counterparts), **Bayesian Classifiers**, and various artificial neural networks. There is no single best approach to modelling content.[2] State of the art systems typically utilize other signals, like user engagement, and thus are discussed in the hybrid system section.

Challenges

Content such as journal text is difficult to use effectively. There is a gap between what people write and what they mean. The team may be in an advantageous position due to text journal being labeled with corresponding user emotions, but still face the challenges of natural language processing.

In addition, there are many aspects of text that affect user preferences, such as readability of a journal. Recent advancements in language models, such as character-level **recursive neural networks** (RNN), have made it easier to adapt to typos, different dialects and a variety of writing styles.[17] However, these models require an immense amount of data, which might call for use of open source models in a process called **transfer learning**.[18,19]

Hybrid Systems

Hybrid systems are the state of the art and combine several data sources to form recommendations.[2] **Knowledge-based** and **demographic-based** signals are typically utilized in hybrid solutions. Knowledge-based approaches utilize explicit knowledge about items or users to build rule-based recommendations. Demographic-based systems

[17] Google Research Blog. Chat smarter with allo, 2016. URL http://googleresearch.blogspot.ca/2016/05/chat-smarter-with-allo.html

[18] Dan C Cireşan, Ueli Meier, and Jürgen Schmidhuber. Transfer learning for latin and chinese characters with deep neural networks. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–6. IEEE, 2012

[19] To use this approach, there still must be enough data to re-train the last layer of a neural network taken from a different machine learning task.

work based on the idea that people with the same demographics, like gender and age, have common preferences.[2]

State of the Art

Hybrid filtering is typically performed using probabilistic methods and ones inspired from biology, including Bayesian networks, neural networks, genetic algorithms,[14] and recently **hopfield networks**.[2]

Collaborative filtering models project users and content into a shared, low-dimensional vector space called a latent space.[20] If a piece of content is close to a user in that space, it would be a good recommendation, given that it has not already been viewed. Interestingly, models can be developed to predict the position of items in that space based on its content.[20] [21] If this position of an item can be predicted accurately, nearby users that would likely enjoy the content can be given a recommendation even if the piece of content has never been seen before. This is a significant leap forward against the cold start problem.

[20] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. Deep content-based music recommendation. In *Advances in Neural Information Processing Systems*, pages 2643–2651, 2013

[21] The implementation used by the author is a deep neural network.

Recommender System Objectives

Users may lose trust if a system is inaccurate. Common ranking quality measures include **precision**, **recall**, normalized **discounted cumulative gain** DCG, and **precision**.[22,2] More recent models evaluate the average of these metrics over the top-N results returned in a recommendation.[2] Such approaches require labeled data, which is expensive to obtain and validate.

[22] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6): 734–749, 2005

Performance Considerations

Often users expect queries to return results rapidly. A two-phase scheme that uses a simple model to select candidates, and a sophisticated one to rank them was proposed.[23] An example of an efficient model is a vector model, such as W2V.

Some features only depend on the item to be ranked, and not its context. For example, journal length and current likes are features that can be pre-computed. Together, they can be utilized to model a quality score, which can speed up ranking.

[23] Andrei Z Broder, David Carmel, Michael Herscovici, Aya Soffer, and Jason Zien. Efficient query evaluation using a two-level retrieval process. In *Proceedings of the twelfth international conference on Information and knowledge management*, pages 426–434. ACM, 2003

Dimensionality Reduction

In both user-generated text and user interactions with content, datasets are **sparse**, where the co-occurrence of any two features is very low. Most users only like a few of the million items in a dataset.

Furthermore, there can be over a million words used across all users, while only a hundred words are used in a single journal. **Dimensionality reduction** is a critical technique for compressing sparse, high-dimensional datasets. It seeks to maintain certain properties in a dataset, such as distances between points, in order to preserve as much information as possible.

Optimization-Based Dimensionality Reduction

A desired property in dimensionality reduction techniques is for the distances between points to be similar in the original space and the reduced space. The sum of differences between points in each space can be computed. The difference can be built in many ways, such as squares or the **Sammon Mapping**, which emphasizes preserving the distances between nearby points over further points.[24]

[24] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9 (2579-2605):85, 2008

After defining a metric in each respective space (it does not necessarily need to be the same metric[24]), one initializes positions by sampling from a normal distribution. Utilizing a stochastic optimization technique like **stochastic gradient descent** (with **momentum**, **normalization** and **temperature control**), one can traverse solution spaces to an approximate optimality. **Temperature** control allows for broader exploration in early optimization, and slow reduction of solution spaces until shortlisted solutions are solved to local optimality. It is related to **simulated annealing**, which has a similar effect but involves more parameter tuning.[24]

t-distributed Stochastic Neighbour Embedding (t-SNE)

Very high dimensional data can be meaningfully visualized to expose clusters through use of **t-SNE**. Each point is connected to a set number of surrounding points which are closest to it, effectively capturing local structure and ignoring other data.[24]

The method visualizes high-dimensional data by mapping each point in space to 2D or 3D. t-SNE uses **random walks** on neighbourhood graphs to allow the implicit structure of all of the data to influence the way in which a subset of the data is displayed. Other non-parametric visualization techniques include **Isomap** and **Locally Linear Embedding**.[24]

Neural Networks

Autoeconders take a non-linear, neural-network based approach to dimensionality reduction. A good introduction to autoencoders is provided by Andrew Ng, a word-class data scientist.[25] In summary, an autoenconder works by modelling inputs X, mapping them to

[25] Andrew Ng. Sparse autoencoders. URL `"web.stanford.edu/class/cs294a/sparseAutoencoder_2011new.pdf"`

a hidden network of lower (or higher, but sparse) dimensionality, and attempting to perform classification on the final hidden layer to predict the original input. In the middle, since dimensionality changes, the model learns interesting patterns within the data, which are revealed by each neuron's activation.

Another model which clusters data naturally is the **self-organizing map** (SOM). It works by creating a network of nodes which interact by bending locally during parameter updating through a **neighbourhood function**. The nodes of an SOM can be clustered.[26]

[26] Juha Vesanto and Esa Alhoniemi. Clustering of the self-organizing map. *Neural Networks, IEEE Transactions on*, 11 (3):586–600, 2000

Principal Components Analysis (PCA)

Principal Components Analysis is the primary linear method employed in dimensionality reduction. A linear mapping to a lower-dimensional space is performed while maximizing variance in that space.[27] This is accomplished by calculating the **co-variance** or correlation matrix of the data, and extracting **eigenvectors** for this matrix. The principal components (vectors corresponding to the largest eigenvalues), thus can be utilized to reconstruct the most variance in the data. The first few eigenvectors thus accomplish dimensionality reduction.

[27] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014

Linear Discriminant Analysis (LDA) is very similar to PCA, but utilizes class variables. For example, different types may exhibit different structure. Clustering can be used to develop latent classes.

Random projections

Taking **random projections** is a simple concept that is efficient. It reduces the dimensionality of data and carefully trades error for faster processing.[28] The data is projected to a much lower subspace using matrix multiplication by a random matrix. Both the dimension and distribution in a random projection matrix are specified to keep distances similar between pairs of items.

[28] Xiaoli Zhang Fern and Carla E Brodley. Random projection for high dimensional data clustering: A cluster ensemble approach. In *ICML*, volume 3, pages 186–193, 2003

Word2Vec

In the case of text mining, the Word2Vec model invented by Google researchers is especially valuable.[29] There are two approaches to this: **Skip-gram** (SG) and **Bag of Words** (BoW). **Negative-sampling** is used to avoid summation over all items in a soft-max layer of a neural network during optimization. The technique provides implicit dimensionality reduction by using only neighbouring words. The output is a vector space with interesting properties, such as allowing for addition/subtraction and clustering of words.

[29] Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. *arXiv preprint arXiv:1405.4053*, 2014

The Future of Recommender Systems

Classic recommender systems utilized user preference histories, demographics, and the content of items to form recommendations. Second-generation engines are utilizing an established social network, such as friend groups, to take advantage of social trust and develop more accurate models for connections between users.[2]

Recent research has pointed to using contextual data about users to build better recommender systems. In different contexts, such as at different times of the day, users may be interested in different types of recommendations.[30] Furthermore, the user's location[31], health signals, eating habits, and even the weather can be useful cues. In particular, it would be novel to classify and predict user moods, and use this as a signal for context.[2] These third-generation systems will be a part of the Internet of Things with various integrated devices such as smart watches, sensors and monitoring services. This makes it an exciting time to be modelling a recommender system based on contextual data about users.

[30] Joseph A Konstan and John Riedl. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2): 101–123, 2012

[31] Location can be used to derive the type of activity a user is engaging in, the news around that area, and which other users are nearby.

Relevant Topics in Mental Health

Mental Health

For this project, it is important to distinguish between mental health and mental wellness. Mental health is defined as the state of emotional psychological well-being that allows an individual to understand and interact with their community productively and to cope with slight variations of stress.[32] Mental health fluctuates over time and poor mental health typically refers individuals whose are easily excited, upset, angered or anxious or have difficulty coping or functioning in regular, everyday situations. When these changes in behaviour occur for an extended period of time, it is possible that the individual is experiencing a mental illness.[33] Although mental illnesses are believed to arise from many different biological and environmental sources, recent research has established a link between poor mental health and both physical and mental illness.[34] Mental illnesses are diagnosed by physicians and there are currently several forms of treatment available.

There is a current push in Canada, North America and Europe to make mental health and mental illness resources more visible. Nearly all of the English speaking countries in these regions have resources available on federal and provincial websites.[35,33] Recent reports made by Statistics Canada show that mental health and wellness affects 1 in 5 Canadians directly and nearly every Canadian has some

[32] What is mental health and mental illness?, 2016. URL http://wmhp.cmhaontario.ca/workplace-mental-health-core-concepts-issues/what-is-mental-health-and-mental-illness

[33] University of Toronto. Mental health vs. mental illness, 2016. URL https://utoronto.morefeetontheground.ca/intro/defining

[34] Government of Canada. About mental health, 2015. URL http://healthycanadians.gc.ca/healthy-living-vie-saine/mental-health-sante-mentale/improving-mental-health-ameliorer-sante-mentale/what-quoi-eng.php

[35] Aware. What is depression?, 2016. URL http://www.aware.ie/help/information/what-is-depression/

relationship with an individual suffering from poor mental health or a mental illness.[31]

Diagnosis and Treatment

Research in mental health and mental illness is still rapidly evolving to define new forms of mental illnesses and new treatments methods. The treatment applied is based on the physician's assessment of the individual's severity of symptoms and the cause (if determined). Current treatments usually comprise of medications and therapy.[34] Mental illness is often a lifelong affliction which can be managed with effective treatment but never completely goes away.[31]

In comparison to mental illnesses, the majority of the recommended treatments for poor mental health are handled by the individuals themselves instead of physicians. The Canadian Mental Health Association provides individuals with "Mental Fitness Tips" on their website to promote good mental health. These tips include regular exercise and sleep, setting personal goals and utilizing journaling or daydreaming to collect and release emotions of the day.[36] The Government of Canada has similar advice as well as a list of common triggers for poor mental health. These include traumatic or stressful life events and toxic relationships or environments. For individuals affected by one or more of these triggers, there are many counselling groups or classes to help individuals learn to accept and overcome their negative or inhibiting emotions.[33] Nearly all of the resources available highlight that having a strong support community is critical to encouraging good mental health.

[36] Canadian Mental Health Association. Mental fitness tips, 2016. URL http://www.cmha.ca/mental_health/mental-fitness-tips

Social Media and Mental Health

For the purpose of this project, it was also important for the group to understand the positive and negative effects of social media use and mental health. In the papers surveyed, social media is defined as online platforms to support social communities and provide tools for sharing knowledge, experiences and opinions.[37] Microblogs, discussed previously, are a form of social media as well as Facebook and video sharing websites like Youtube. Since social media is an emerging technology, research in the field is limited to this millennium. Additionally, studying the long term effects of social media on its users like their physical and mental health is still evolving. However, it was possible to identify some common opinions on the positive use of social media to encourage communication about mental health and the negative effects social media can have on vulnerable individuals. Specifically, most of the literature on the negative effects of social

[37] Rosemary Thackeray, Benjamin T Crookston, and Joshua H West. Correlates of health-related social media use among adults. *J Med Internet Res*, 15(1): e21, 2013. DOI: 10.2196/jmir.2297

media are related to children and teenagers and individuals with suicidal tendencies.

One of the main benefits of social media platforms is their accessibility. As internet and computer or smartphone access increases, more individuals are able to reach social media sites. This is due to the fact that most social media sites are free to join. Nearly two thirds of adult American internet users have at least one social media account.[36] Thackeray at al. also found that 60% of internet users utilize the internet to search for health related information and 15% utilize online health tracking tools (like Stigma). This reinforces the common theme in most of the literature; social media platforms eliminate geographic and monetary boundaries to make mental health education and support more accessible. Lloyd argues that cyber healthcare has nearly no cost to operate and the tools can be reached immediately.[38] These same benefits are summarized by Grajales et al.[39]

[38] Alfie Lloyd. Social media, help or hindrance: What role does social media play in young people's mental health. *Psychiatria Danubia*, 26(1):340–346, 2014

[39] Francisco Jose Grajales III, Samuel Sheps, Kendall Ho, Helen Novak-Lauscher, and Gunther Eysenbach. Social media: a review and tutorial of applications in medicine and health care. *Journal of medical Internet research*, 16(2):e13, 2014

The second positive feature of social media repeated in many of the literature consulted was that social media allows small, isolated individuals suffering from mental health issues to join online communities dedicated the these issues.[40] Lloyd references a study completed in 2010 where 29% of the individuals interviewed felt it was easier to discuss private or difficult topics online than in person. Other research has shown that the creation of minority communities on Facebook (such as pediatric cancer, ethnic minorities and LGBT youth) empowers many individuals to share their experiences. This in turn educates unaffected individuals.[37]

[40] David D Luxton, Jennifer D June, and Jonathan M Fairall. Social media and suicide: a public health perspective. *American Journal of Public Health*, 102 (S2):S195–S200, 2012

The negative influences of social media on mental health are mostly concentrated on the ability to access pro-suicide content (such as suicide note posted to social media newsfeeds) or the phenomenon known as cyber-bullying.[39] As of the research published so far, it is difficult to prove the causality of social media or the internet as a whole since there are too many offline factors which cannot be modelled. Studies have been able to prove that cyber-bullying, which occurs when an individual is targeted with harassment or threats, can increase poor mental health when the individual is already unstable or depressed. Lloyd argues that cyberbullying can become more intense than offline harassment with the ability to create fake profiles and to publicize the harassment on the platform.[37]

References

What is mental health and mental illness?, 2016. URL http://wmhp.cmhaontario.ca/workplace-mental-health-core-concepts-issues/what-is-mental-health-and-mental-illness.

Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734–749, 2005.

Canadian Mental Health Association. Mental fitness tips, 2016. URL http://www.cmha.ca/mental_health/mental-fitness-tips.

Aware. What is depression?, 2016. URL http://www.aware.ie/help/information/what-is-depression/.

Oren Barkan and Noam Koenigstein. Item2vec: Neural item embedding for collaborative filtering. *arXiv preprint arXiv:1603.04259*, 2016.

Yoshua Bengio, Aaron Courville, and Pierre Vincent. Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1798–1828, 2013.

Daniel Billsus and Michael J Pazzani. Learning collaborative information filters. In *ICML*, volume 98, pages 46–54, 1998.

Google Research Blog. Chat smarter with allo, 2016. URL http://googleresearch.blogspot.ca/2016/05/chat-smarter-with-allo.html.

Jesús Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutiérrez. Recommender systems survey. *Knowledge-Based Systems*, 46:109–132, 2013.

John S Breese, David Heckerman, and Carl Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, pages 43–52. Morgan Kaufmann Publishers Inc., 1998.

Andrei Z Broder, David Carmel, Michael Herscovici, Aya Soffer, and Jason Zien. Efficient query evaluation using a two-level retrieval process. In *Proceedings of the twelfth international conference on Information and knowledge management*, pages 426–434. ACM, 2003.

Dan C Cireşan, Ueli Meier, and Jürgen Schmidhuber. Transfer learning for latin and chinese characters with deep neural networks. In *Neural Networks (IJCNN), The 2012 International Joint Conference on*, pages 1–6. IEEE, 2012.

Xiaoli Zhang Fern and Carla E Brodley. Random projection for high dimensional data clustering: A cluster ensemble approach. In *ICML*, volume 3, pages 186–193, 2003.

Francisco Jose Grajales III, Samuel Sheps, Kendall Ho, Helen Novak-Lauscher, and Gunther Eysenbach. Social media: a review and tutorial of applications in medicine and health care. *Journal of medical Internet research*, 16(2):e13, 2014.

Zan Huang, Wingyan Chung, Thian-Huat Ong, and Hsinchun Chen. A graph-based recommender system for digital library. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, pages 65–73. ACM, 2002.

Joseph A Konstan and John Riedl. Recommender systems: from algorithms to user experience. *User Modeling and User-Adapted Interaction*, 22(1-2):101–123, 2012.

Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.

Quoc V Le and Tomas Mikolov. Distributed representations of sentences and documents. *arXiv preprint arXiv:1405.4053*, 2014.

Alfie Lloyd. Social media, help or hindrance: What role does social media play in young people's mental health. *Psychiatria Danubia*, 26 (1):340–346, 2014.

David D Luxton, Jennifer D June, and Jonathan M Fairall. Social media and suicide: a public health perspective. *American Journal of Public Health*, 102(S2):S195–S200, 2012.

Andrew Ng. Sparse autoencoders. URL "web.stanford.edu/class/cs294a/sparseAutoencoder_2011new.pdf".

Government of Canada. About mental health, 2015. URL http://healthycanadians.gc.ca/healthy-living-vie-saine/mental-health-sante-mentale/improving-mental-health-ameliorer-sante-mentale/what-quoi-eng.php.

University of Toronto. Mental health vs. mental illness, 2016. URL https://utoronto.morefeetontheground.ca/intro/defining.

Michael J Pazzani and Daniel Billsus. Content-based recommendation systems. In *The adaptive web*, pages 325–341. Springer, 2007.

Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798. ACM, 2007.

Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

Rosemary Thackeray, Benjamin T Crookston, and Joshua H West. Correlates of health-related social media use among adults. *J Med Internet Res*, 15(1):e21, 2013. DOI: 10.2196/jmir.2297.

Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. Deep content-based music recommendation. In *Advances in Neural Information Processing Systems*, pages 2643–2651, 2013.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

Juha Vesanto and Esa Alhoniemi. Clustering of the self-organizing map. *Neural Networks, IEEE Transactions on*, 11(3):586–600, 2000.