International Journal of Artificial Intelligence, Internet of Things and Cloud Computing
Volume: 1 (2022), pp. 29-36

A Review on Movie Recommender Model Using Machine Learning

Nandini Mehta[1], Dr. Jyoti Chaudhary[2], Shivani Garg[3]

[1] Research Scholar, Shri JJT University, Jhunjhunu, Rajasthan, 333053, India
[2] Assistant Professor, TIT&S, Bhiwani, Haryana, 127021, India
[3]Assistant Professor, Aravali College of Engineering & Management, Greater Faridabad, Haryana, 121006, India

Abstract

Internet works with clients to convey and share snippets of data. Because of the tremendous measure of data, it has turned into a major test to observe what we need that best meets our inclinations. This issue is alluded to as over-burden. RS have arisen as a Web personalization instrument that helps online clients by lightening the data over-burden issue and permit the disclosure of valuable data. An effective recommender system always ensures that it captures the actual preferences of the users and recommends only those things that the user actually wants. In the last two decades, recommender systems have been implemented in diverse areas of applications to recommend products, contents, and services to online users. Despite the success of recommender systems in various application domains, there are many issues that affect the performance of recommender systems. There are many extensions to recommender systems that can capture the actual preferences of the users by alleviating such issues.

Hence, two levels of filtering are employed to identify similar users; first, during clustering level through multi-criteria ratings, and second at matching level through overall ratings. Several experiments are carried out based on Yahoo! Movies dataset to explain the importance of this model. The results demonstrate that the performance of clustering based recommender systems is better than the commonly used similarity measure based recommender systems on the basis of various evaluation measures.

Keywords: Recommendation System; Deep Learning; Filtering; AI; Rating; Precision.

1. Introduction

Recommender systems are being implemented in diverse areas of applications to recommend products, contents, and services to internet users. Various online platforms join recommender systems to extend shopper services, grow selling percentage, and reduce users searching time. To validate the impact of Recommender systems in different domains, McKinsey explored that around 35% of the sales on Amazon are made because of the recommender systems. Also, Alibaba achieved a growth of up to 20% during the Chinese global shopping festival in November 2016. On YouTube and Netflix, around 60% and 80% of the time people spend watching videos comes from recommendations, respectively. Furthermore, Google News generates 38% more click-through recommendation. These organizations have successfully adopted commercial recommender systems and have increased Web sales with enhanced users' involvement and satisfaction. Recommendation system categorized into various application areas such as:

• Entertainment: In this field, RS have increased tremendous prevalence as well as demonstrated to be an immense achievement. From movies and music to jokes, recommender systems have left a never-ending sway. MovieLens.com, Amazon Prime, Netflix, YouTube are a couple of examples of recommender systems developed for entertainment. For instance, on MovieLens.com, when we watch and rate a movie highly from some specific genre, we get suggestions for different movies from a similar type of genre.

• E-learning: These systems provide recommendation for documents, personalized newspapers, e-learning applications, recommendation of Web pages, or blog recommendation. E-learning has picked up its acceptance in educational institutes for longer than ten years at this point. Such platforms aim to provide study materials to the users depending on their leanings and past exercises. Digital libraries or e-libraries are the base of e-learning where the users can discover various information and knowledge sources. Coursera and edx are notable examples of this category.

• E-commerce: Over the years, various recommender systems have been implemented in the field of Web-based businesses to provide recommendation for internet users of products to purchase, for example, CDs, books, digital cameras, laptops, and so forth. User feedback and past preferences form the mutual criteria for making the recommendation. Several shopping sites have been developed, such as Amazon, Flipkart, and Alibaba, etc. On these sites, we generally observe recommendation as top-selling, similar products, most viewed, people also purchased, etc.

• Social networking: Social networking sites are too useful these days. These sites keep individuals connected and inform the users what their friends, relatives, and partners are doing. Facebook, LinkedIn, Twitter and so on are renowned examples. Recommendation in such platforms can be found as pages you may like, people you may know, suggestions for you, etc. Here recommendations are usually produced using browsing patterns, clicking percentage, and through user tags.

• Services: recommendation of e-government services, travel services, experts for consultation services, financial services, hotel, and restaurant recommendation are the examples of this category. For instance, e-tourism gives recommendation to the tourists about spots worth visiting when making a trip to a specific location. For more comfort and ease, such systems have presently evolved as mobile apps and can be accessed anywhere by the user. TripAdvisor, Goibibo, Booking.com are some of the examples of this category.

• Editorial and hand curated: Editorial and hand curated items include a list of favorites and a list of essential items. These type of recommendation generally depends upon the interest of a user on very limited products. For example, staff picks, home pages of websites, etc. Editorial and hand curated items don’t scale very well and require a lot of human participation and fine-tuning.

• Simple aggregates: The simple aggregates or non-personalized technique is the one that does not consider the personal interest of the users to understand whether they can accept the items or not. In other words, the same recommendation is given to every user in a non-personalized recommendation—for example, top 10, most popular, most recent YouTube videos. The predicted ratings are completely neither dependent on the users’ purchasing behavior nor their matching with other users. They usually use the aggregates method to obtain the collective scores given to a particular item and assign the value to every user who did not rate the item. Thus, these simple aggregates don’t take users’ personalization into account.

• Tailored to individual users: Contrary to simple aggregates recommendation, tailored recommendation uses various techniques to estimate personal opinions of the users and make appropriate recommendation for users that might be specifically useful to them. Tailored recommendations are also known as personalized recommendation and most of the e-commerce sites provide tailored recommendation to individual users such as Amazon, Netflix, and LinkedIn recommendation systems.

2. Types of Recommendation System

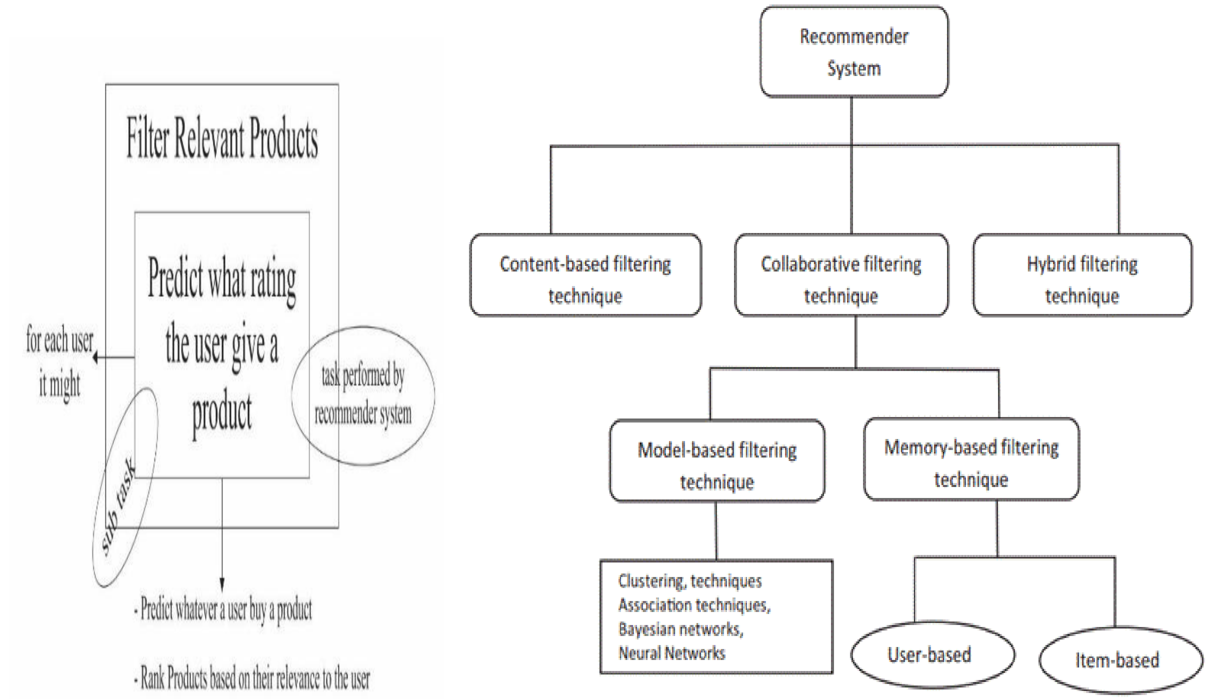

Fig1. (a) Filter Relevant Products; (b) Recommender System

The most conspicuous procedures of the suggestion frameworks are the Content and Collaborative separating strategy, Hybrid proposal, Knowledge-based sifting, Demographic technique and Model-based method. A few specialists utilize both the blend of these strategies for suggestion frameworks. In the substance based separating, the crucial interaction depends on purchaserportrayals and their required thing. Then, at that point, a profile is overseen addressing the things through meaning it to the objective client who recently loved something similar. The Collaborative separating technique is the most eminent strategy likewise broadly utilized in items, administrations and travel suggestions. This is likewise a typical strategy for planning the proposal framework. It utilizes a monstrous volume of information gathered from the way of behaving of the client in a previous time and predicts which thing the clients like most.

Data mining along with machine learning are the progressive approach that is commonly a part of suggestion techniques. In such groups, the regulations of the approach are to be learned while optimizing frame chart. Few checks under this category will consist – role of latent factor, rule-based proposals, Bayesian methods and decision trees. Considering the rating scenario where count is less, models with latent factor play a crucial role. Hybrid Filtering is when many algorithms, strategies, and concepts used in collaborative and content wise filtering are merged to produce a definitive positive output. Product attributes, customer preferences, ratings, comments, and ideas are all combined here to forecast a specific behavior and ensure accuracy throughout the process. To create such a system, we can also state that it is more complex and involves deep machine learning techniques.

Table1. Deep Machine Learning Technique

RS Technique	Advantage	Non-advantage
Collaborative	No information designing exertion, high possibility of results, learns market portions	Needs few type of rating criticism, cold beginning for new clients and new things
Content-based	No people group known, the examination between things conceivable.	Content portrayals vital, cold beginning for new clients, no curve balls

Knowledge based	Checking suggestions, guaranteed service, no cool beginning, can look like deals discourse	Information designing work to bootstrap, fundamentally static, doesn't respond to momentary patterns
Hybrid Approach	Adaptability of various constraint, both content and review based analysis.	Needs intense back up for data density issue, step by step procedure required for the analysis.

3. Literature Review

Koyel Dutta et al [1], focused on significant approach in filtering and accessing the online information in a suitable way. Terminologies such as Utility matrix and long tail are given a brief overview and these are considered crucial to understand the approach of recommender system. Also, various hybrid approaches have been discussed and a successful attempt was made in this direction. The paper aims to give a proper insight of recommendation system and based on content and various hybrid techniques. New and basic items that are always in trend in the market such as books, restaurant, music etc. are taken as examples to verify the role of this system. A new book that gets launched consists of reviews, demand and selling energy before it gets in the hands of the buyers. Now a day's people prefer to go to a restaurant based on previous visit by other consumers and their reviews posted on the social media including staff nature, ambience, services offered and food quality. Such examples have eased in understanding the concept of recommendation system. This system helps in improving the decision-making process of an individual who are eager to gain knowledge and want to access the best services offered.

M.P. Geetha et al [2],describes utilization of deep learning techniques in recommendation system to increase the feature of predicted output. E commerce sells all kinds of products which makes it highly complex to make a choice. Users are suggested with an option to choose the best based on the previous user interaction with the product. Data acquisition phase, learning phase, feedback phase is explained to understand the hierarchy of the model presented. Also, the following research deals in deep investigation of data and how recommendation system proves to be an effective tool in accomplishing the future preferences of a consumer. Data on a daily basis is generated in a large amount in GBs due to which them exist limitations that are discussed in coming section. To tackle and overcome these problems, deep learning which is a sub-domain of machine learning is utilized in this system. The following paper deals with the fundamental concepts of the recommendation engine including all the phases. Further feedback gathered is investigated and an appropriate algorithm is applied that reflects the output of recommendation system.

Najafabadi et al., [3]in his research suggested deep learning to be the quickest aspect of machine learning to abstract the true sense of data and further to understand different levels of data representations it uses deep neural networks. This whole concept is attracting entrepreneurs and their firms to work in accordance with the output of these studies and research work. Here, algorithms used in deep learning help us to better understand recommendation system.

The techniques used here can be applied to both supervised and unsupervised learning methods. DBN and AE refers to Deep Belief Networks and Auto Encoders are procedures subjected to unsupervised group. Also, RBM which is Restricted Boltzmann Machines and DBM meaning Deep Boltzmann Machines fall in the same category. GAN is General Adversarial Network is the latest concept in this learning method.

RNN, CNN and MLP referring to Recurrent Neural Network, Convolutional Neural network and Multi-Layer Perception are some of the learning methods for supervised category.

Batmaz et al., [4]suggested that content-based filtering relates to the information available regarding user and product service. This is the prime and main concern in developing a standard approach in the recommender system. Text messages, images, videos, audio can influence a user to depict his interest in the product marketing model. Various surveys are conducted to observe the behavior and reaction of the consumers towards a service.

How this is done? To get to the final output, the system generates a profile of consumer's preferences based on available samples of products with their content information. Also, textual and visual information of the user is extracted for recommendation purpose.

Zhang et al., [5]observed resolution strategy in the field of digital applications. This was done through collaborative

Filtering. Here, a matrix based on user-item relation was evaluated and then predictions were made on this basis. User ratings were collaborated with the neighbors views to come up with a positive impact. For this approach, the ratings provided by users were normalized and a collective prediction was calculated.

F. Strub et al., [6] suggested Auto Encoder which is an Artificial Neural Network to collect information from a set example or input data in the form of text, image or audio. It comprises of several functions and it is a three-layer network. To understand the concept involved here, a function gets registered with the input data and this gives an output. The following output is used as an input and transferred to the hidden layer in a compressed data representation.

One thing to be aware of is that we need same number of neutrons for input and output layer as at last we have to rebuild and gain the original data.

The suggested model consists of an Encoder which diminishes the input data in reduced dimensions and further provide it to the hidden layer.

The other one is the Decoder which converts data at hidden layer to its original form and creates output. The hidden layer is the most important as it consists of a rating matrix which helps to construct a recommendation system.

Su X et al., [7] illustrated RBM, Restricted Boltzmann Machine which is a graphical model and has equal importance in deep learning framework as other techniques. It has a couple of neuron layers in which unsupervised data is provided to the input layer and then, data features are acquired by the hidden layer. Such model helps in dimensionality reduction of the data and further encourages the Deep Belief Network concept.

Y. K. Tan et al., [8] followed the concept of RNN which is Recurrent Neural Network to provide suggestions in development of recommendation system. Sequential data is generated by understanding the reviews made by a user according to his preferences. It catches the data populating from stock markets, government agencies whether it is in handwriting form, words spoken most often or numeric series. The process uses Back Propagation through Time (BPTT) model which gives a clear view on the suggested approach.

User holds a session in which he checks for a couple of products, short list them and waits for the correct time to purchase it. The session is analyzed using this model and predictions are made.

Da'u, A et al., [9] propagated the work under GAN, General Adversarial Network. It consists of generative and discriminative neural networks.

In the generative network, a query is related with the relevant items whereas in the discriminative network, it is complete opposite.

GAN is an effective way to distinguish the target data formed in generative network with the training data produced in discriminative network.

Zhang et al., [10] developed a two-stage collaborative filtering Recommender System, especially for diversified products. The first stage of the RS relates multiple interests with neighbor selection to obtain the complete and diversified product list. In the next stage considers the social relationships of the user to predict ratings by available ratings. The trust ranking mechanism is used to rearrange the list of TOP-N recommended items with specified threshold values. A broad experimental valuation considered in a real-world dataset, the proposed model achieves higher diversity and accuracy compared to traditional methods.

4. Problems On Hand

Most of the Conventional Recommendation system is prone to various drawbacks due to ineffective behavior. Some issues are as follows:

Table1. Deep Machine Learning Technique

Problems & Challenges in RS	Description	Impact/ Risk Ratio
Consumers Inter-relation	The mentioned defect is based on the association of the members like X is similar to Y, Y is similar to Z than in a way Z will be equalizing X. This equation is hard to determine by the classifiers.	High-Low
Cold-start	It is in situation where there is a little information	Low –High

	about content is available and information without user ratings. This issue affects every recommendation system.	
Sparsity	This limitation relates with the shopping nature of the customer, he/she rates/purchase a few products out of the wide range of available goods via online commerce websites.	High-Low
Definition	This term relates to defect when products are nearly equal as per title. Example – Windcheater and Sweatshirt are different and are part of a same class 'Clothing'.	Neutral
Privacy	Privacy issue was found in most of the techniques. Since the personalization factor is not addressed when the information is processed in bulk.	High-Low
Standard	A classifier needs to maintain the quality of work by its output and it should be neglecting the false negatives. Example, A user liking a specific brand 'Puma' but the other user B has rated it as bad, upon using, A finds it extremely useful, so, this situation creates a negative impact which should be clarified by the classifier.	Low-High
Scalability	More and more users having access to the web will consequently start purchasing online. This increase the graph of reviews/ratings inserted and hence, it now becomes essential to develop efficient recomm-ender classifier.	Intermediate

5. Procedure

Commitment strategy is used to ascertain client conduct and recognize boundaries for online films, for example, normal watch time which is executed with film length property and video quality. Late examination has shown that personalization proposal is subject to the nature of a computerized thing connected to the client's choice to keep watching or tuning in. Consequently, they want to evaluate the time given by the user while seeing the movie and also the quality substance.

The module takes input from the client history as contribution with the UHC module. The primary goal of this stage is to take the contribution from key ID of motion pictures being watched and delivers yields that could for the most part be proper to the client inquiries and wants. It recommends results investigating client inclinations and working with them to deal with the data over-burden issue. CRM - Candidate Recommendation Module structure comprehends the necessities of the clients and presents ramifications of a few cinematographic things. This data can be as the past utilization of things or the evaluations that were given to the things. CRM approach looks for things or motion pictures that are like the ones clients have watched or have preferred before. The past suggestion frameworks basically chipped away at content separating measures (CFM).

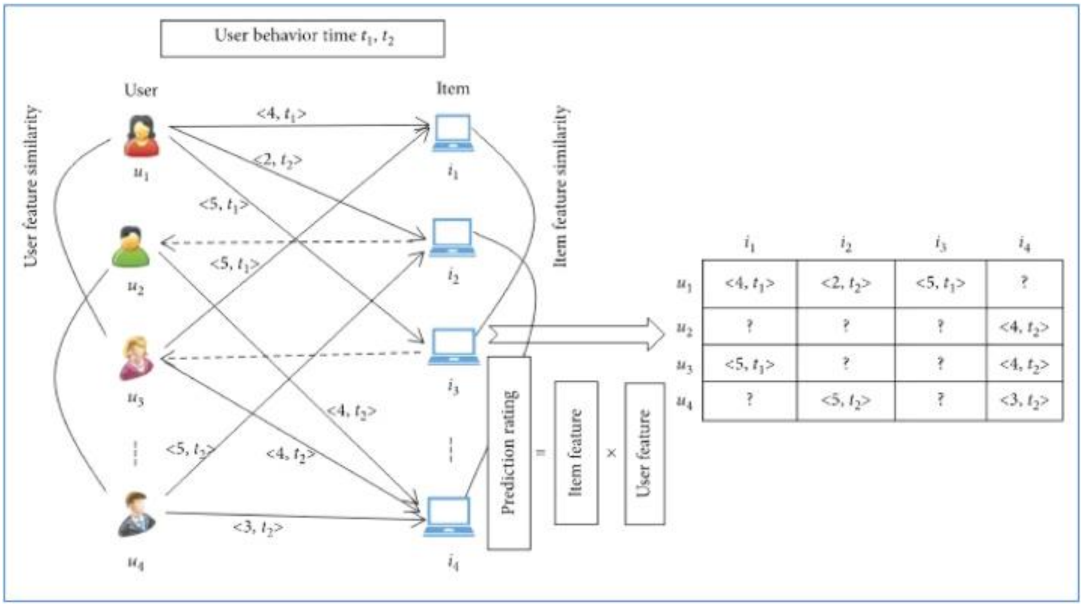

Fig3 User Feature Similarity

6. Result

This segment portrayed the examinations and their outcomes during testing of the model. It likewise results a few related techniques to begin investigation on existing works and our suggested work to precisely lay out possible benefits of proposed structure. In particular, test examination plans on the advancement of personalized suggestions in the application region.

6.1 Dataset

Here, Movie Lens dataset has been considered and partitioned into test information. The investigations are executed on two unique releases of the dataset. In the trial and test information, all information comprises of 20 arbitrary datasets with those clients who appraised something like 25 motion pictures. The information data of hundred clients is utilized for the trial reason, 150 clients in the test information and 1000 evaluations from 60 clients on 90 films. The client's segment data is utilized for the client's recognizable pieces of proof like age, orientation, and movement. The appraisals decide from 0 to 5 rating scale. All clients have characterized their data with an ID and other data is portrayed.

6.2 Evaluation

While carrying out such model, either a clever calculation or an inventive methodology is expected to realize how well the structure executes. The most ordinarily used investigation strategies in the proposed framework are Predictive Accuracy Metrics and mean equal position (MRR), and this large number of methods are utilized for the valuation of the proposed system execution. To assess the presentation of a structure, things' forecast is executed by using the proposed system and not entirely settled by contrasting the anticipated and real appraisals of the things.

To assess the exactness of the structure the essential still up in the air from the quantity of things which are either applicable or superfluous? These sets can be unequivocally characterized in a possibility table lattice. Accuracy and review are the main predominant measures for the assessment of the proposed system. Accuracy is depicted as the quantity of related things chose to the quantity of things liked and it estimates the productivity of applicable things liked. In accuracy, the TPA means genuine positive precision that deliberate as the proportion of prescribed things which are connected with the whole rundown.

Precision=TPA=TP upon/ TP+FP

Table2. Evaluation

	Relevant	Irrelevant	Total
Recommended	TP	FP	TP+FP
Not Recommended	FN.	TN	FN+TN
Total	TP+FN	FP+TN	N

FP = false positive.

FN = false negative.

TP = true positive

TPR = true positive rate

7. Conclusion

In this paper, we proposed an individual suggestion framework to deal with related films relying on perspectives and preferences. Since suggestions are for the most part utilized for a considerable length of time targets, for example, client fulfillment or improved exchanges, estimating ultimately requires considering this and assessing the deliberate outcome. Nonetheless, it very well may be costly to endeavor approaches on genuine gatherings of clients and work out the results. Our structure helps clients to find data that best accommodates their cravings and inclinations in an over-burden search space and has been centered at the precision improvement. Later on, we are anticipating applying this model to a huge informational collection.

References

1. KoyalDatta Gupta, 2019, “A Survey on Recommender System”, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 14, Number 14 (2019) pp. 3274-3277
2. M.P.Geetha, D.KarthikaRenuka, 2019, “Research on Recommendation Systems using Deep Learning Models”, International Journal of Recent Technology and Engineering (IJRTE)ISSN: 2277-3878, Volume-8 Issue-4, November 2019
3. Najafabadi, Maryam &Villanustre, Flavio&Khoshgoftaar, Taghi&Seliya, Naeem& Wald, Randall &Muharemagic, Edin. (2015). Deep learning applications and challenges in big data analytics. Journal of Big Data. 2. 10.1186/s40537-014-0007-7.
4. Batmaz, Zeynep&Yürekli, Ali & Bilge, Alper&Kaleli, Cihan. (2018). A review on deep learning for recommender systems: challenges and remedies. Artificial Intelligence Review. 10.1007/s10462-018-9654-y.
5. Zhang F,YuanNJ, Lian D, XieX,MaWY(2016a) Collaborative knowledge base embedding for recommender systems. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, San Francisco, CA, USA, pp 353–362.
6. F. Strub and J. Mary, ``Collaborative filtering with stacked denoisingautoencoders and sparse inputs," in Proc. NIPS Workshop Mach. Learn.e Commerce, Montreal, QC, Canada, 2015, pp. 150_158.
7. Su X, Khoshgoftaar TM (2009) A survey of collaborative filtering techniques. Advances in Artificial Intelligence, 2009(Section 3), 1–19. https ://doi.org/10.1155/2009/42142 5.
8. Y. K. Tan, X. Xu, and Y. Liu, ``Improved recurrent neural networks for session-based recommendations," in Proc. 1st Workshop Deep Learn. Recommender Syst., Boston, MA, USA, 2016, pp. 17_22.
9. Da’u, A., & Salim, N. (2019). Recommendation system based on deep learning methods: a systematic review and new directions. Artificial Intelligence Review, 1 – 40.
10. Zhang F,YuanNJ, Lian D, XieX,MaWY(2016a) Collaborative knowledge base embedding for recommender systems. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, ACM, San Francisco, CA, USA, pp 353–362.