Application of ML industries 2

Computer Science (Anna University)

Welcome to:

Machine Learning in Communication, Media and Entertainment

Unit objectives

After completing this unit, you should be able to understand:

- Understand the purpose of using machine learning in communication, media and entertainment

- learn about the usage of machine learning in media and entertainment industry

- Learn about analyse the customer sentiment using machine learning models

- Understand need of real-time analytics in communication, media, entertainment industries

- Gain knowledge on deep learning for social media analytics

- Learn about different types of recommendation engines

- Understand the Restricted Boltzmann Machines (RBM) for collaborative filtering

- Gain knowledge on collaborative deep learning for recommender systems

Machine learning in communication, media and entertainment

- Identification and conviction of media content helps the customers to communicate easily with new media and materials from various sources successfully with the help of machine learning and data science techniques.

- AL and ML are the main technologies in the telecommunication industry that assist the firms to generate better income, build more trust from the customer end and have good customer relationships.

Usage of machine learning in media and entertainment Industry

- The contributions of analytics in the entertainment space are:
  - Helps us to understand consumer insights/psychology.
  - Tracking the customer's digital footprint to schedule ad campaigns accordingly.
  - Enhancing the product based on customer feedback.
  - To creating content based past available data.

- The following are the ways by which data science and machine learning are changing the entertainment and media industry:
  - Prediction of audience behaviour.
  - Analysing customer sentiment
  - Personalization of content

Machine learning techniques for customer sentiment analysis

- The concept of applying natural language processing and text analysis techniques to recognize and draw out subjective information from a piece of text is called sentiment analysis.

- Majority of the part, the feelings or opinions of a person are subjective and not actual facts.

- Another name given to sentiment Analysis is opinion mining, which is an area within Natural Language Processing (NLP) which builds up systems that can recognize and withdraw opinions within text.

- Word embedding is the collective name for a set of language modelling and feature learning techniques in Natural Language Processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers.

- Conceptually it involves a mathematical embedding from a space with many dimensions per word to a continuous vector space with a much lower dimension.

- Methods to generate this mapping include neural networks, dimensionality reduction on the word co-occurrence matrix, probabilistic models, explainable knowledge base method, and explicit representation in terms of the context in which words appear.

Sentiment analysis with long short term memory networks

- LSTM networks are a type of RNN that uses special units in addition to standard units.

- LSTM units include a 'memory cell' that can maintain information in memory for long periods of time.

- A set of gates is used to control when information enters the memory, when it's output, and when it's forgotten. This architecture lets them learn longer-term dependencies.
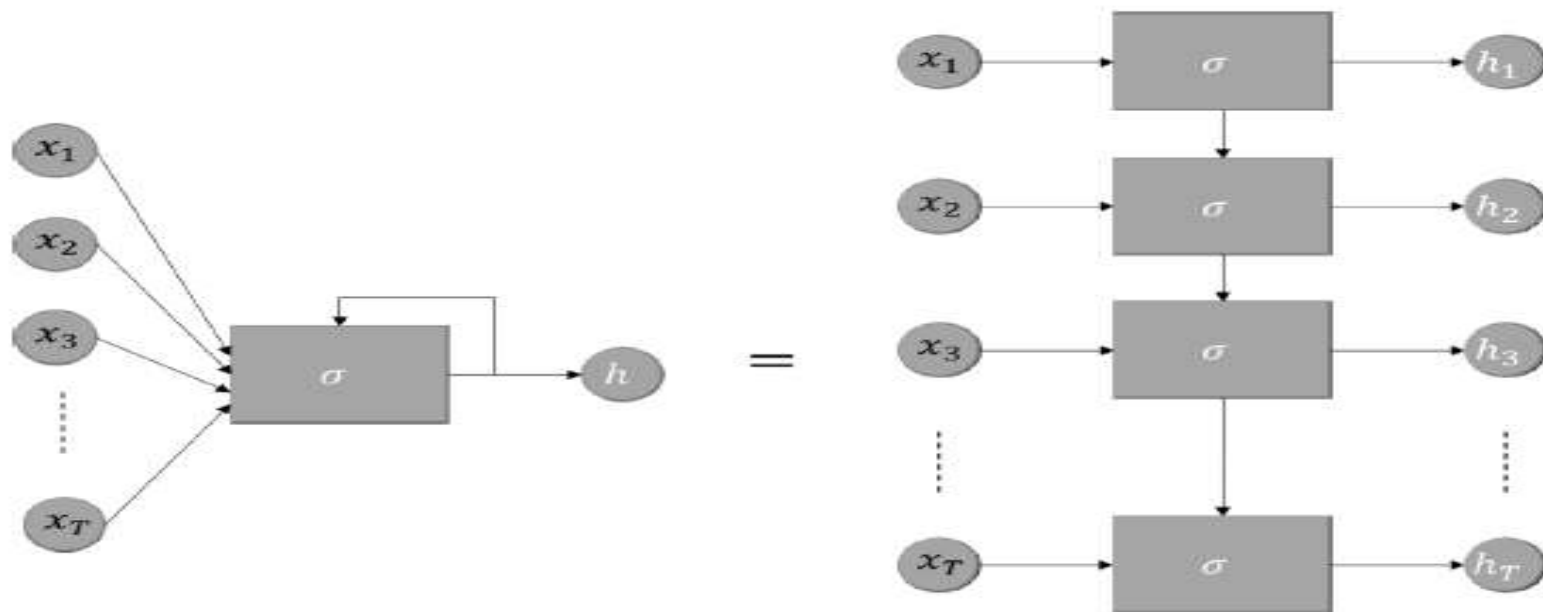

Figure: Recurrent neural network

Source: https://www.math.vu.nl/~sbhulai/papers/paper-edema.pdf

- Real time analytic is used in analysis of data as soon as the data become available.

- As soon as the data enters the system the customers were able to draw conclusions.

- Social sites like Facebook, Twitter and Instagram were used to advertise their products and services to brand it across the globe, media customers and entertainment customers had intensively high competition.

- The areas in which social media analytics make a big impact are as follows:
  - Innovation.
  - Marketing.
  - Sales.
  - Costumer services.
  - Competitive ontelligent.

Deep learning for social media analytics

- Social media analytics is the practice of gathering data from social media websites and analysing that data using social media analytics tools to make business decisions.

- The most common use of social media analytics is to mine customer sentiment to support marketing and customer service activities.

- The content generated by the user are composed by interactive web 2.0 internet-based applications such as:
    - Texts.
    - Posts.
    - Comments.
    - Videos.

- All the data is generated through online communications.

Recommendations engines

- The recommender system technology recommend users and help setting up in the environment for things like:

    - Products.

    - Movies.

    - Events.

    - Articles.

- Types of recommendation engines:

    - Collaborative filtering.

    - Content based filtering.

    - Hybrid recommendation systems.

- This filtering technique is based on the resemblances of the users and its built on collecting and analyzing information of user's activities such as their:

  - Behaviors.
  - Choices.

- This predicts what the user may like or prefer.

- Types of collaborative filtering:

  - Memory based collaborative filtering.
  - Content based collaborative filtering.

- Advantages of memory based collaborative filtering are:

  – Easier implementation in any condition.

  – Easier update of database can be done so as the new arrival data can easily be picked up.

  – Large sized data can be processed.

  – Feedback is provided to explain the working process of the recommender system by providing evidence such as listing the characteristics or illustration which made a product to appear in the suggestion site.

- Limitations of memory based collaborative filtering techniques are:

  – If active users have anything in contact with the individual who classified the product to be suggested, then the predictions made by the memory-based CF techniques were inaccurate and unreliable.

  – It makes prediction slow as it uses entire database each time.

Memory based collaborative filtering (2 of 2)

- Equation 1:

$$W_{u,v} = \frac{\sum_{i \in I}(r_{u,i} - \overline{r_u})(r_{v,i} - \overline{r_v})}{\sqrt{\sum_{i \in I}(r_{u,i} - \overline{r_u})^2}\sqrt{\sum_{i \in I}(r_{v,i} - \overline{r_v})^2}}$$

- Equation 2:

$$W_{i,j} = \frac{\sum_{u \in U}(r_{u,j} - \overline{r_i})(r_{u,j} - \overline{r_j})}{\sqrt{\sum_{u \in U}(r_{u,j} - \overline{r_i})^2}\sqrt{\sum_{u \in U}(r_{u,j} - \overline{r_j})^2}},$$

- Equation 3:

$$W_{i,j} = \cos(\vec{i}, \vec{j}) = \frac{\vec{i} \cdot \vec{j}}{\|\vec{i}\| * \|\vec{j}\|}$$

- **Advantages:**

  - Over fittings can be avoided easily and the dataset is scalable.

  - Improvement in prediction performance.

- **The limitations of model-based CF algorithms are:**

  - Due to inflexibility it becomes hard to attach information to model based systems for users who do not rate.

  - Since it's not able to generate reasonable recommendations it suffers from sparsity problems.

- The profile of the user's choices and description of an item are the inputs for content-based filtering.

- The user profile is created to understand the sort of product an individual likes, keywords are also used to describe the item.

- If we like a product, we have the tendency to like a similar item and that's the principle behind content-based filtering.

- Hybrid methods can be implemented by independently creating and creating and integrating content based and collective predictions or in addition, by applying a cooperative approach to a content on the basis of capability and vice versa, or by integrating methods into a single model.

- The various ways of implementation of hybrid approaches are:
  - Implement content-based and collaborative methods separately and accumulate their observation.
  - Some of collaborative aspects are comprised into the content based approach.
  - Some content based features are integrated into collaborative approach.
  - General consolidate model is constructed which is integration of both collaborative and content based characteristics.
  - The problems faced in recommendation systems such as cold start and sparsity is resolved in hybrid recommender technique.
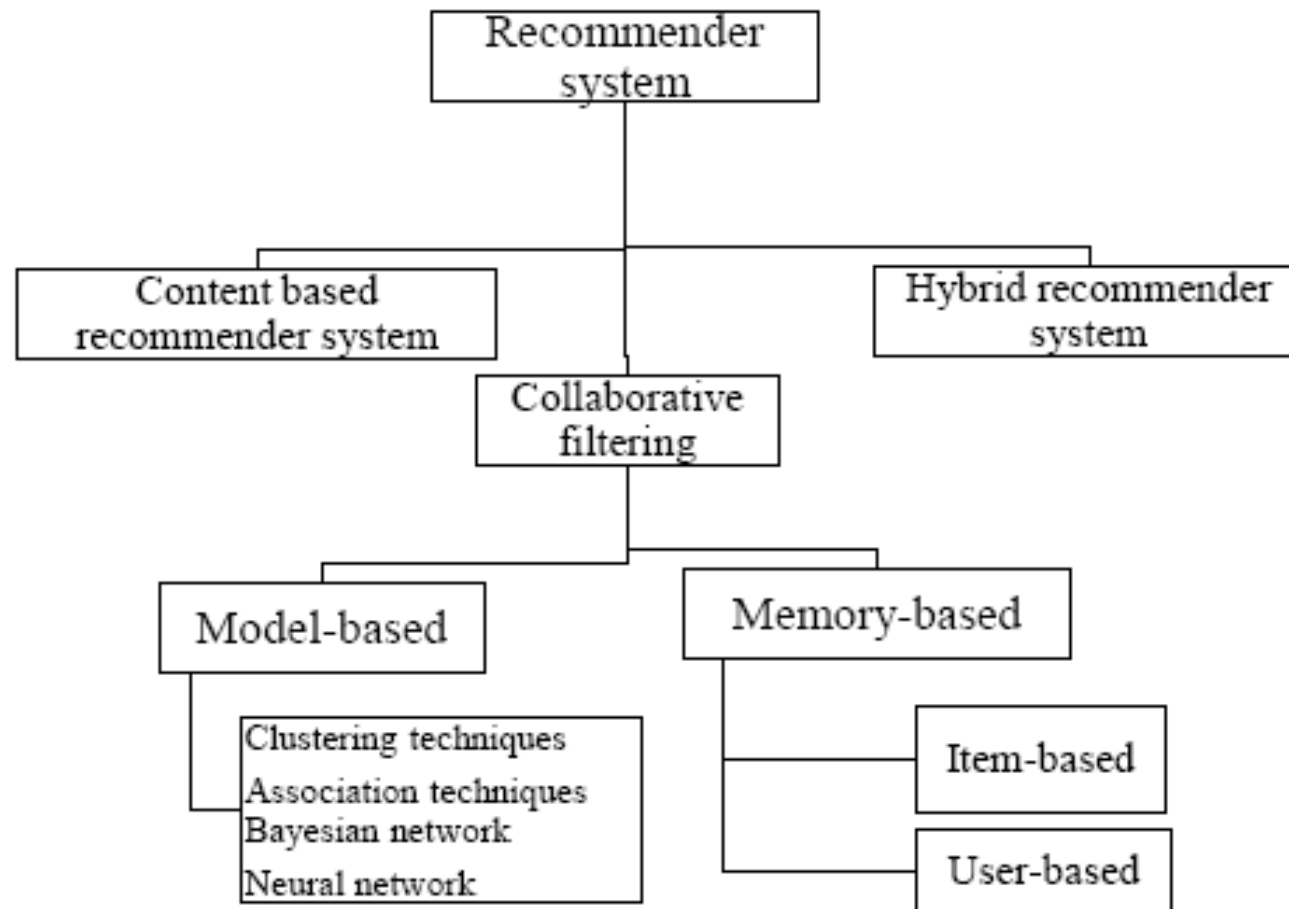
Summary of recommendation systems

Figure: Summary of recommendation systems
Source: https://www.math.vu.nl/~sbhulai/papers/paper-miedema.pdf

Deep learning techniques on recommender systems

- Building recommender systems for collaborative and content-based approaches, deep learning has been suggested.

- Restricted Boltzmann Machines (RBM) for collaborative filtering:
  - The special version of Boltzmann Machine (BM) consists of layer hidden components and a layer of visible components with no hidden-hidden or visible-visible contacts is the Restricted Boltzmann Machine (RBM).

- Collaborative deep learning for recommender systems:
  - Collaborative deep learning for recommender systems was introduced to address the cold start problems and it utilizes review texts and ratings.

- Bayesian Stack De-Noise Auto Encoder (SDAE) and Collaborative Topic Regression (CTR) is integrated to collaborative deep learning.

Multiple choice questions:

1. Word embedding is the collective name for a set of language modelling and feature learning techniques in Natural Language Processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers.

   a) Vector

   b) Scalar

   c) Letter embedding

   d) Word embedding

2. Which analytics is used in analysis of data as soon as the data become available?

   a) Full time analytics

   b) Real time analytics

   c) Time analytics

   d) Media analytics

3. Which of the following is a type of recommendation engine?

   a) Content based filtering

   b) Sentence based filtering

   c) Word based filtering

   d) Collective based filtering

Multiple choice questions:

1. Word embedding is the collective name for a set of language modelling and feature learning techniques in natural language processing (NLP) where words or phrases from the vocabulary are mapped to vectors of real numbers.

    a) Vector

    b) Scalar

    c) Letter embedding

    d) Word embedding

2. Which analytics is used in analysis of data as soon as the data become available?

    a) Full time analytics

    b) Real time analytics

    c) Time analytics

    d) Media analytics

3. Which of the following is a type of recommendation engine?

    a) Content based filtering

    b) Sentence based filtering

    c) Word based filtering

    d) Collective based filtering

Fill in the blanks:

1. --------------- is a sub-field of NLP that tries to identify and extract opinions within a given text across blogs, reviews, social media, forums, news etc.,

2. --------------- is the collective name for a set of language modelling and feature learning techniques in natural language processing where words or phrases from the vocabulary are mapped to vectors of real numbers.

3. ---------------- is used in analysis of data as soon as the data become available.

4. ---------------- technique is based on the resemblances of the users and its built on collecting and analyzing information of user's activities such as their behaviors or choices.

Fill in the blanks:

1. **Sentiment analysis** is a sub-field of NLP that tries to identify and extract opinions within a given text across blogs, reviews, social media, forums, news etc.,

2. **Word embedding** is the collective name for a set of language modelling and feature learning techniques in natural language processing where words or phrases from the vocabulary are mapped to vectors of real numbers.

3. **Real time analytic** is used in analysis of data as soon as the data become available.

4. **Collaborative filtering** technique is based on the resemblances of the users and its built on collecting and analyzing information of user's activities such as their behaviors or choices.

Two mark questions:

1. Define word embedding's.
2. What is real-time analytics?
3. What is social media analytics?
4. What is sentiment analysis?

Four mark questions:

1. Explain the basics of sentiment analysis.
2. List the usage of machine learning in media and entertainment industry.
3. Explain the deep learning techniques on recommender systems.
4. Explain hybrid recommendation systems.

Eight mark questions:

1. Write a short notes on sentiment analysis with long short-term memory networks.
2. Need of real-time analytics in communication, media and entertainment industries.

Having completed this unit, you should be able to:

- Understand the purpose of using machine learning in communication, media and entertainment

- learn about the usage of machine learning in media and entertainment industry

- Learn about analyse the customer sentiment using machine learning models

- Understand need of real-time analytics in communication, media, entertainment industries

- Gain knowledge on deep learning for social media analytics

- Learn about different types of recommendation engines

- Understand the Restricted Boltzmann Machines (RBM) for collaborative filtering

- Gain knowledge on collaborative deep learning for recommender systems