UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

DIPARTIMENTO DI INFORMATICA

PhD Program in Computer Science and Mathematics
XXXIV **cycle**

Research Project

PhD Student: Andrea Iovine

Supervisor: Dott. Marco de Gemmis

Coordinator: Prof.ssa Maria F. Costabile

PhD student signature Andrea Iovine

Supervisor signature Marco de Gemmis

1. Research title:
Conversational agents for recommender systems

2. Research area:
Natural Language Processing, Conversational Systems, Information Retrieval, Information Filtering

3. Research motivation and objectives
Since the introduction of the World Wide Web and, subsequently, the mass diffusion of social networks, people have been given access to an unprecedented amount of information. This has caused the so-called *information overload* phenomenon: since humans have limited cognitive processing capabilities, such a quantity of information negatively impacts their decision-making abilities, by making it more difficult to make effective decisions (Jameson et al. [1])**.**
In order to support the decision-making process, Recommender Systems have been developed. They are software tools that can suggest products or services in which the user might be interested, based on a user profile [2].

Recommender Systems methodologies can be roughly classified as:

- **Collaborative Filtering**, which suggest items from users with similar interests to the current user;
- **Content-Based Filtering**, which recommend items that are similar to the ones liked by the user, based on intrinsic features of the items themselves.

Canonical recommender systems are evaluated by their ability to recommend relevant items, however that does not mean necessarily that those items are useful to the user, e.g. when a recommended item is already known by the user. For this reason, new techniques and metrics have been explored in order to create and evaluate recommender systems that are more useful for the user.
The need for a more user-oriented recommendation framework means that researchers are moving away from classical, query-based Information Retrieval models. In some situations, in fact, a user may not be able to express their requirements in query form or may not know at all what they are looking for. The *information seeking* tasks conducted by the users require several steps, such as *sensemaking* and *exploratory search*, which in turn require collecting, organizing and understanding large quantities of information in order to make a decision. Information seeking is also an interactive process, as it requires multiple search sessions in order to be completed [3,4].

The objective of this project is to develop new types of Information Seeking Support Systems which exploit Conversational Agents to elicit semantic features from multiple sources. The distinctive characteristic of the approach is the use of Conversational Agents that will support the system by providing an interaction mode that mimics the interactive information seeking process followed by the users. In order to reach this objective, some challenges need to be addressed, which are:

- The development of techniques for extracting knowledge from text;
- The development of automatic user requirements extraction methods;
- The development of techniques for exploiting semantic features for producing useful recommendations;
- The design and development of methods for handling the context of the conversation and using it in order to understand the user's goal;
- The development of Conversational Systems that can be easily applied to different domains and scaled to manage a high number of user actions, contexts, and goals.

4. State of the art

Classic Information Retrieval metrics such as Precision, Recall and F-measure can only determine if a Recommender System is able to give relevant recommendations; however they give no information on how useful the recommendations are to the user. For this reason, both new techniques and new metrics have been explored in order to create and evaluate recommender systems that are more useful for the user. de Gemmis et al. [5] conducted an investigation on the subject of *serendipity*, which is usually defined as the experience of unexpected and fortuitous recommendations. A Recommender System is able to give *serendipitous* recommendations if it suggests unexpected yet interesting items to the user. Serendipity can be used both as a solution for the overspecialization problem faced by Content-Based Recommender Systems, and as a general principle for developing systems that are able to give actually useful recommendations to the user.

Pearl Pu, et al. [6] developed ResQue (Recommender Systems' Quality of User Experience), a framework for the evaluation of Recommender Systems in a user-centric perspective. ResQue evaluates not only how much the recommendations match the interests of the users, but also evaluates how much the users believe that the system is easy to use, how much they are satisfied by using the system, and how much they trust the suggestions given by the system.

An interesting approach to develop Recommender Systems that are more useful and that produce a better user experience is to introduce a conversational interface. The resulting systems are called *Conversational Recommender Systems* (CoRS). In a CoRS the user can interact with the system by exchanging messages, usually through natural language. Conversational Recommender Systems divide the interaction process into iterative phases: first they acquire information from the users (for example, their preferences), they recommend some items based on that information, and then they receive a feedback from the user (also sometimes called *critiquing*), and use that feedback to generate a better set of recommendations, until they recommend an item that is of interest for the user.

An example of critiquing-based conversational recommender system is ReComment, presented in Grasch et al. [7], that uses a speech-based interface to communicate with the user. In each interaction step a product is presented, and the user can provide a critique (e.g "*cheaper*" or "*different manufacturer*").

Recently, the inclusion of user requirements within the recommendation process is being investigated. Indeed, some users may ignore or not be interested in the technical features of a product, but rather, they want to know what's the best product for their specific tasks. Bogers and Koolen [8] define the concept of *Narrative-driven Recommendation* (NDR), a new paradigm in which the recommendation process is not only driven by past transactions, but also by a narrative description of their current interests. Intuitively, the concept of NDR is very relevant for CoRS, since a narrative description is a very good source of user requirements.

In 2014, Widyantoro et al. [9] presented a conversational recommender system framework, which is able to provide recommendations and explanations based on the user's functional requirements. The user requirement extraction task is conducted via a question-answering dialog. The system combines *navigation-by-asking* and *navigation-by-proposing*: during the interaction it asks questions about some functional requirements. When the system has enough information, it generates a list of recommended items. As for the semantic level, both functional requirements and product features are handled by using an ontology.

In 2015, Arnett et al. [10] presented a recommender system which extracts user functional requirements and product technical features. The Euclidean Fuzzy similarity metric is used to calculate the mapping between the user's requirements and the features of a product. The system has been tested in the smartphone recommendation scenario. Sharma et al. [11] refined this solution by introducing a trapezoidal fuzzy function.

Christakopoulou et al. [12] developed a preference elicitation framework that supports conversational recommender systems by identifying which questions should be asked to a new user in order to quickly learn their preferences. The question selection task is conducted via *bandit learning*, while

the recommendation uses a Collaborative Filtering approach based on *Probabilistic Matrix Factorization* (PMF).
In 2018, Sun and Zhang [13] developed a conversational recommender system based on faceted search, which allows users to narrow down a list of products by adding constraints on a group of facets. The system is implemented as a text-based chatbot, which uses a hybrid End-to-End architecture for belief tracking, dialog policy and recommendation. It also uses *deep reinforcement learning* in order to select the next facet to ask. The belief tracker, based on LSTM neural networks, extracts facet-value pairs from the user messages. The recommendation network is based on a factorization machine. The policy network uses reinforcement learning and selects the next action.

Since providing useful recommendations requires deep understanding of the semantics behind both user requirements and products, researchers are investigating the use of machine learning techniques for extracting meaning from data. One of the most popular approaches is the use of *aspect-based sentiment analysis* (or *opinion mining*) from user reviews, which can be used to discover aspects that are of interest for the user, along with the sentiment orientation for each aspect. Aspect-based sentiment analysis for recommender systems was introduced in Zhang et al. [14], which uses *latent Dirichlet allocation* (LDA) to extract aspects from the reviews, and SentiWordNet[1] in order to associate a sentiment score for each aspect. Both users and products are then represented as vectors: a product vector p_i is extracted from all the reviews for that product, while a user vector u_j is extracted from the user's reviews. The recommendation score can then be calculated as a similarity between p_i and u_j.
The *Sentiment Utility Logistic Model* (SULM), presented by Bauman et al. [15], also uses aspect-based opinion mining from user reviews for enhancing the recommendation process. This information is used to predict items, along with the most important aspects that may enhance the user experience. The opinion mining component not only extracts the aspect-sentiment pairs from a review, but also evaluates the impact of each aspect on the overall rating. The recommendation model is then trained to predict the user's rating of the aspects of an unknown item. The advantage compared to traditional recommendation models is that the aspects are automatically extracted, and thus they are not constrained to a predefined set. The system has been tested in restaurant, hotel, and beauty&spa applications.

In 2017, Musto et al. [16] developed a multi-criteria recommender system based on Collaborative Filtering, which exploits user reviews to generate a multi-faceted user model. It uses an opinion mining and sentiment analysis framework in order to extract aspect, sentiments, and the relations between aspects and sub-aspects, using the *Kullback-Leibler divergence* measure. The recommendation is then generated using a Multi-Criteria User-to-User Collaborative Filtering approach.
A different approach for using reviews is described in Zhang et al. [17], which introduces the *System Ask - User Respond* (SAUR) paradigm for conversational search, and an implementation of said paradigm using *Personalized Multi-Memory Networks* (PMMN). In the SAUR paradigm the user initiates the task by providing an initial request, to which the system responds by asking a series of faceted questions. Once the system has enough confidence, it can provide a list of recommendations that the user can either accept or reject. The network is trained using review data in order to identify the right order of the questions. Each review is transformed into a simulated conversation.
Recently, Rafailidis and Manolopoulos [18] discussed the idea of putting together Virtual Assistants and Recommender Systems, highlighting the gap between the two technologies. The authors sustain that a conversational assistant can improve the recommendation process because it has the ability to learn the users’ evolving, diverse and multi-aspect preferences.

[1] https://sentiwordnet.isti.cnr.it/

Aside from the recommendation aspect, much research is being conducted in the area of dialog systems. In order to go beyond the recommendation task, a conversational information seeking support system must be able to handle vast, multi-domain, goal-oriented dialogs. *End-to-End dialog systems* are a new achievement in the area, which use end-to-end deep learning in order to learn a conversation model directly from dialog examples, without any need for ad-hoc components. Results are very promising; however there is a strong need for a goal-oriented framework, for evaluating whether the agent is able to direct the conversation towards the user's goal. Dodge et al. [19] presents a dataset for the evaluation of End-to-End systems in movie recommendation and question answering tasks.

Williams et al. [20] presents a model for end-to-end learning of task-oriented dialog systems, based on Hybrid Code Networks. This model is able to combine RNN with explicit domain knowledge expressed via software and action templates. This allows the system to maintain a latent representation of the dialog state while specifying behavior that is hard to learn but easy to program (i.e. in a banking dialog system, requiring that the user is logged in before performing any transaction). The proposed system can be trained both via supervised and reinforcement learning.

An End-to-End system based on neural networks is proposed in Liu et al. [21]. It can be used to develop task-oriented conversational systems. The system is composed of a single neural network that is able to do Dialog State Tracking, call an API, and generate a response. In this case the dialog is modelled as a multi-task sequence learning problem, using a bi-directional LSTM neural network. The system has been compared to other neural network-based solutions, using the Dialog State Tracking Challenge 2 dataset, resulting in better performance compared to the competitors.

5. Problem approach

In order to find a solution to the challenges mentioned in Section 3, we will investigate the following approaches:

- Aspect-based Opinion mining techniques for the extraction of user requirements and product features from review data. Reviews are an invaluable source of information about the users' experience towards a product, because they can highlight the products' perceived qualities. An effective conversational recommender should be able to recognize the needs of the user and find the most appropriate items that fit those needs. In particular, review data can be exploited:
    - to identify functional requirements of the active user, based on their own reviews, by discovering the aspects of an item that are most relevant for them through opinion mining techniques;
    - to enrich the feature set of an item in a Content-Based Recommender System, by adding subjective evaluations to the static set of objective features. In this case, opinion mining techniques can be applied to all the reviews for a product, in order to identify the most relevant aspects and corresponding sentiment of users on them.
- Explanation techniques, which are an integral part of the previously mentioned sensemaking process. An effective explanation mechanism should help users to take better decisions, making the system more trustworthy [22]. Techniques such as feature-based explanation will be investigated.
- End-to-end Deep Learning for the development of goal-oriented conversational systems. By training the agent directly on conversational data, it will be able to handle even complex conversations, that can keep track of a great number of user goals, that can be scaled and easily moved to new domains.
- Techniques for combining structured and unstructured knowledge sources to best fit he user needs.

The application of the aforementioned approaches requires the development and the execution of a plan, which includes the following steps:

1. Acquisition of item data, review data and dialog data from various sources. Items and reviews can be usually found in e-commerce services or social networks. Conversational data can be retrieved from real users, or from state-of-the-art datasets.
2. Application of an aspect-based opinion mining framework on the review data, in order to extract the opinion of users.
3. Development of a recommender algorithm that can handle aspect-based recommendations.
4. Training of the recommender model with the results of Step 2, and training of the conversational interface with the dialog dataset.
5. Evaluation of the effectiveness of recommendations and the quality of the conversation.

6. Expected results

The approaches that will be investigated in this project have multiple applications. Industry is particularly receptive to the idea of applying conversational interfaces and aspect-based opinion mining to Information Seeking Support Systems. The domains of interest include:

- E-commerce: Service providers are very interested in exploiting the reviews stored in their systems in order to improve their recommendation process. Having a more useful, insightful and trustworthy recommender system means that users are willing to make more purchases. At the same time, it will be easier for customers to find the items that best fit their requirements, by reducing the time needed for the intensive information seeking task. That goal can be also achieved by infusing in the conversational agent knowledge derived from structured knowledge sources (e.g. data coming from catalogs).
- Customer Support and Relationship Management: Conversational agents can provide support to customers, by reducing the need for human operators. Also, both recommendation and customer support can be integrated into one system, which becomes the unique point of contact between the business and the customer.
- Product Analysis: manufacturers or service providers can use aspect-based sentiment analysis of the reviews to gain insight about what users think about their products or services, as well as to identify the most relevant and critical product features.
- Domains in which a completely hands-free interaction is needed. A conversational agent can help users find the information they need even when doing intensive tasks that demand minimal distraction. An example of such domain is the automotive one, where digital assistants such as Mercedes-Benz's *MBUX*[2] are currently being developed, which feature a completely voice-based interface. A conversational agent can replace many of the hand-operated controls of a car, reducing distraction and thus increasing safety.

7. Phases of the project

1st year: study of literature state of the art and basic research material

Activity 1.1: study of the state of the art regarding Goal-based Conversational Agents and Information Seeking Support Systems.
Activity 1.2: Investigation and analysis of the techniques for Aspect-Based sentiment analysis, conversational recommenders and explanation techniques.
Activity 1.3: Attendance of courses and seminars in accordance with the study plan.
Activity 1.4: Participation to international schools and conferences that are relevant for the goals of this research project.

[2] https://www.daimler.com/innovation/case/connectivity/mbux-2.html

2nd year: development of methods

> **Activity 2.1:** Gap analysis: study of the possible improvements over the approaches developed by other researchers.
> **Activity 2.2:** Development of techniques for Opinion Mining, Conversational Recommendation and Explanation, and acquisition of the datasets for training and evaluating them.
> **Activity 2.3:** Evaluation of the developed solution, comparison with existing state of the art, and publication of the results.

3rd year: Application and writing of the Ph.D thesis

> **Activity 3.1:** Collaboration and comparison with national and international research groups, with possible internships in foreign universities or research groups.
> **Activity 3.2:** Application of the developed methods to the aforementioned domains, and analysis of the results.
> **Activity 3.3:** Preparation of the Ph.D thesis.

The following table shows the mapping between each activity and the time period in which it will be carried out.

	1st year				2nd year				3rd year			
Activity	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**
1.1												
1.2												
1.3												
1.4												
2.1												
2,2												
2.3												
3.1												
3.2												
3.3												

8. Result evaluation

Evaluation of the results of the techniques described in this project requires both *in-vitro* and *in-vivo* experiments, each with its own metrics. In-vitro experiments will be conducted using state-of-the-art datasets, while in-vivo studies will require setting up a user study. The metrics can be divided into three categories: classical IR metrics, user-oriented recommendation metrics, and explanation metrics.

The classical IR metrics are:

- **Hit Rate:** $HR@n = \frac{\#hits}{n}$, which is the number of correct recommendations (hits) in an *n*-sized ranked list of recommendations.
- **Mean Average Precision:** $MAP@k$ is the mean of the Average Precision calculated over all recommendation lists. The Average Precision is:

$$AP@k = \frac{\sum P@l \cdot rel(l)}{k}.$$

$P@l$ is the precision calculated on the first l elements of the list, and $rel(l)$ is a function that returns *1* if the user liked the item in position *l*, *0* otherwise.

- **normalized Discounted Cumulative Gain:**

$$nDCG@k = \frac{DCG@k}{IDCG@k}$$

where the Discounted Cumulative Gain $DCG@k$ is calculated as:

$$DCG@k = \frac{1}{|U|}\sum_{u\in U}\left(r_{ui_1} + \sum_{j=2}^{\|R_u\|}\frac{r_{ui_j}}{\log_2(j)}\right),$$

in which U is the set of users, R_u is the k-sized vector of ratings associated to the user u, r_{ui_j} is the rating of user u to item i_j. $IDCG@k$ is the Ideal Discounted Cumulative Gain, which is calculated as $DCG@k$ in which the R_u vectors have been sorted by rating in descending order.

In order to capture the users' objective opinion, the preparation of user studies will be necessary. Data can be collected and analyzed via the submission of a questionnaire. An example of such questionnaire is the ResQue model, described in [6].
The effectiveness and the quality of the conversational interaction can be evaluated objectively via metrics such as the Task Success Ratio, which calculates the percentage of conversations in which the user found the information they needed. Another measure is the number of dialog turns that were dedicated to recovering anomalous situations (e.g. when the system does not understand the user). Subjective evaluations are also possible via the analysis of questionnaire data, such as the *SASSI* framework [23].
The effectiveness of an explanation mechanism [22] can be evaluated by comparing the user rating of an item before and after the explanation has been given, or by comparing the user's agreement with the explanation before and after the consumption of said item.

9. Possible reference persons external to the department

In this section are identified people from national or international Universities or research groups, who represent an authority on the research areas of this project, and with whom I hope to establish a communication and a collaborative effort.

1. Prof. Marco Gori, Dipartimento Ingegneria dell'informazione e scienze matematiche, Università di Siena, Italy
2. Prof. Francesco Ricci, Facoltà di Scienze e Tecnologie informatiche, Libera Università di Bolzano, Italy
3. Prof. Oliver Lemon, Department of Computer Science, Heriot-Watt University, Edinburgh, United Kingdom

10. References

[1] A. Jameson, M. C. Willemsen, A. Felfernig, M. de Gemmis, P. Lops, G. Semeraro, and L. Chen, "Human decision making and recommender systems," in Recommender Systems Handbook, pp. 611–648, Springer, 2015.

[2] F. Ricci, L. Rokach, and B. Shapira, "Introduction to recommender systems handbook," in *Recommender systems handbook*, pp. 1–35, Springer, 2011.

[3] D. Kelly, S. Dumais, and J. O. Pedersen, "Evaluation challenges and directions for information-seeking support systems," *Computer*, vol. 42, no. 3, pp. 60–66, 2009.

[4] P. Pirolli, "Powers of 10: Modeling complex information-seeking systems at multiple scales," *Computer*, vol. 42, no. 3, pp. 33–40, 2009.

[5] M. de Gemmis, P. Lops, G. Semeraro, and C. Musto, "An investigation on the serendipity problem in recommender systems," *Information Processing & Management*, vol. 51, no. 5,pp. 695–717, 2015.

[6] P. Pu, L. Chen, and R. Hu, "A user-centric evaluation framework for recommender systems," in *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys '11, (New York, NY, USA), pp. 157–164, ACM, 2011.

[7] P. Grasch, A. Felfernig, and F. Reinfrank, "Recomment: towards critiquing-based recommendation with speech interaction," in *Proceedings of the 7th ACM Conference on Recommender Systems*, pp. 157–164, ACM, 2013.

[8] T. Bogers and M. Koolen, "Defining and supporting narrative-driven recommendation," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 238–242, ACM, 2017.

[9] D. H. Widyantoro and Z. Baizal, "A framework of conversational recommender system based on user functional requirements," in *Information and Communication Technology (ICoICT), 2014 2nd International Conference on*, pp. 160–165, IEEE, 2014.

[10] D. K. Arnett, Z. Baizal, et al., "Recommender system based on user functional requirements using euclidean fuzzy," in *Information and Communication Technology (ICoICT), 2015 3rd International Conference on*, pp. 455–460, IEEE, 2015.

[11] S. Sharma, C. R. Krishna, S. Solanki, K. Kaur, and S. Kaur, "User-interactive recommender system for electronic products using fuzzy numbers," in *Advanced Computing and Communication Technologies*, pp. 73–81, Springer, 2018.

[12] K. Christakopoulou, F. Radlinski, and K. Hofmann, "Towards conversational recommender systems," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 815–824, ACM, 2016.

[13] Y. Sun and Y. Zhang, "Conversational recommender system," *arXiv preprint arXiv:1806.03277*, 2018.

[14] Y. Zhang, R. Liu, and A. Li, "A novel approach to recommender system based on aspect-level sentiment analysis," in *2015 4th National Conference on Electrical, Electronics and Computer Engineering*, Atlantis Press, 2015.

[15] K. Bauman, B. Liu, and A. Tuzhilin, "Aspect based recommendations: Recommending items with the most valuable aspects based on user reviews," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 717–725, ACM, 2017.

[16] C. Musto, M. de Gemmis, G. Semeraro, and P. Lops, "A multi-criteria recommender system exploiting aspect-based sentiment analysis of users' reviews," in *Proceedings of the Eleventh ACM Conference on Recommender Systems*, pp. 321-325, ACM, 2017.

[17] Y. Zhang, X. Chen, Q. Ai, L. Yang, and W. B. Croft, “Towards conversational search and recommendation: System ask, user respond,” in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, pp. 177-186, ACM, 2018.

[18] D. Rafailidis, “The technological gap between virtual assistants and recommendation systems”, *arXiv preprint arXiv:1901.00431*, 2018

[19] J. Dodge, “Evaluating prerequisite qualities for learning end-to-end dialog systems,” 2016.

[20] J. D. Williams, K. Asadi, and G. Zweig, “Hybrid code networks: practical and efficient end-to-end dialog control with supervised and reinforcement learning,” *arXiv preprint arXiv:1702.03274*, 2017.7

[21] B. Liu and I. Lane, “An end-to-end trainable neural network model with belief tracking for task-oriented dialog,” *arXiv preprint arXiv:1708.05956*, 2017.

[22] N. Tintarev and J. Masthoff, “Evaluating the effectiveness of explanations for recommender systems,” *User Modeling and User-Adapted Interaction*, vol. 22, no. 4-5, pp. 399–439, 2012.

[23] K. S. Hone and R. Graham, “Towards a tool for the subjective assessment of speech system interfaces (sassi),” *Natural Language Engineering*, vol. 6, no. 3-4, pp. 287–303, 2000.