1. Title : Unsupervised detection of chorus waves in spectrogram data.

Description :Chorus waves are a special kind of coherent electromagnetic emissions found in the magnetospheres of planets like Earth and Mercury. Till date, there is no unified theory on its properties, so building a deterministic detector is hard. Researchers have begun characterizing visually which signatures could belong to chorus waves, and still a lot of work needs to be done. In this regard, Deep learning and specifically representation disentanglement methods may prove useful in detecting possible samples. Since data is abundant, but labels are scarce, an active learning paradigm could also be applied to manually label the most important samples. A catalogue prepared in such a manner could be beneficial to planetary scientists to map out the characteristics of such waves.
Supervisors : Sahib Julka : Sahib.Julka@uni-passau.de (Prof. Michael Granitzer Michael.Granitzer@Uni-Passau.De, Uni Passau)
—---

2. Title : Evaluation of properties in disentanglement representation learning.

Description: Disentanglement representation learning deals with factorising latent spaces into deterministic and interpretable sub spaces. Ideally these representations need to satisfy some properties such as modularity, completeness, group symmetry and mutual invariance. Currently there is no consensus on which properties are absolutely essential and which satisfy varying definitions of disentanglement. In this thesis, we will analyse the existing proposed properties, and evaluate them in regards to different definitions and metrics. We will clarify these properties based on task agnostic, specific and investigate how the properties can be learnt with and w/o the use of inductive biases.
Supervisors : Sahib Julka : Sahib.Julka@uni-passau.de (Prof. Michael Granitzer Michael.Granitzer@Uni-Passau.De, Uni Passau)
—---

3. Title : Generating vector representations of Kernel data in Linux Systems

Description: An OS Kernel is the heart of an operating system and contains all the necessary processes to manage the hardware and software resources. We are looking to create a fixed-length vector representation for the OS Kernel to obtain information about the kernel and the different structures and processes running within the kernel. In addition, the aim is to produce a generalized representation for kernels(Mem2Vec) to cluster the kernels based on their version and possibly detect if malicious programs are running in the kernel.

Supervisors : Christofer Fellicious: Christofer.Fellicious@uni-passau.de (Prof. Michael Granitzer Michael.Granitzer@Uni-Passau.De, Uni Passau)

----

4. Title : Generating sequence aware representations for API Call Stack

Description: Application Programming Interface(API) is a method of requesting some functionality or data from a program. An example of an API call can be a request to open a file in the file system. Programs use APIs to carry out tasks such as sending or receiving data from the web, downloading content or evening writing/reading data to the disk. Malicious programs mostly have a specific sequence of execution of such API calls. Malware programs cloak these sequences within random junk API calls to hide their specific signatures. The length of such API call sequences can vary from a few tens to thousands of API calls. Therefore, we need to generate a sequence-aware representation of the API calls to classify malware into different families and identify future malware based on the previously learned signatures.

Supervisors : Christofer Fellicious: Christofer.Fellicious@uni-passau.de (Prof. Michael Granitzer Michael.Granitzer@Uni-Passau.De, Uni Passau)

----

5. Title : Food healthiness scores in the context of a Food Recommender

Description :

This topic is in the field of food recommender systems. The goal of such a system is to recommend food related items (like recipes) to a given user. Traditional recommender systems in this domain only tend to take users' food preferences (e.g., a user likes spicy food) into account. However, a "health-aware" system should also focus on medical aspects. Thus, it is of utmost importance to include nutritional information into the system.

For each recommended food item, it is important to know details about micro- & macro nutrients, and how to rate those for a given user. There are scores defined by official health organisations like the WHO **[Amine2003]**, FSA **[FoodStandardsAgency2016]**, or EFSA **[Efsa2022]** which must be implemented and adapted for this use-case. Related works as **[Trattner2017, Trattner2017a]** could be a good starting point as they use their custom adapted metrics to evaluate food items on their healthiness.

On the other side, for each user, it is essential to have medical insights as for example potential allergies or targeted diets. The goal is to formalize those requirements in order to match those with potential recommendations (e.g., limit all possible recommendations by possible ones or filter a list of recommended items by invalid ones).

Although this topic is heavily related to a food recommender in the machine learning (ML) domain, the topic can be carried out without any ML knowledge.

The research questions are stated as follows:

1. How can official food scores or traffic light systems be adapted to measure the "healthiness" of a food item like a recipe?

2. Can those measures be adapted to measure the "healthiness" of a food item with a given restriction of a certain diet?

3. How can those scores be used to measure the "healthiness" of a whole dataset (e.g., all the recipes from "food.com")?

4. Which are the key insights when measuring the items of the custom dataset?

The evaluation can be processed on the custom dataset crawled from "food.com" which includes recipes, nutritional information, and user-item interactions as reviews and ratings.

References

[Trattner2017] Trattner, C., & Elsweiler, D. (2017). Food Recommender Systems: Important Contributions, Challenges and Future Research Directions. Retrieved from http://arxiv.org/abs/1711.02760

[Trattner2017a] Trattner, C., & Elsweiler, D. (2017). Investigating the healthiness of internet-sourced recipes implications for meal planning and recommender systems. In 26th International World Wide Web Conference, WWW 2017 (pp. 489–498). International World Wide Web Conferences Steering Committee. https://doi.org/10.1145/3038912.3052573

[Amine2003] Amine, E. K., Baba, N. H., Belhadj, M., Deurenberg-Yap, M., Djazayery, A., Forrestre, T., … Yach, D. (2003). Diet, nutrition and the prevention of chronic diseases. World Health Organization - Technical Report Series. https://doi.org/10.1093/ajcn/60.4.644a

[FoodStandardsAgency2016] Food Standards Agency. (2016). Guide to creating a front of pack (FoP) nutrition label for pre-packed products sold through retail outlets. Food Standards Agency, (November), 33. https://www.food.gov.uk/sites/default/files/media/document/fop-guidance_0.pdf

[Efsa2022] The European Food Safety Authority (2022). Dietary Reference Values. https://www.efsa.europa.eu/en/topics/topic/dietary-reference-values

Supervisors : Felix Bölz, Diana Nurbakova, Sylvie Calabretto

—---

6. Title : A probabilistic meta-model for inference detection on both personal and sensor data

Description :
https://www.overleaf.com/read/mggnwgvhmptw

Supervisors : Paul Lachat, Nadia Bennani , Veronika Sonigo
—---

7. Title : Transforming Data into Linked Data and Improving Performance in the context of healthy food recommendation

Description :
This topic is located in the domain of food recommender systems. The goal of such a system is to recommend food related items (like recipes) to a given user. Traditional recommender systems in this domain only tend to take users' food preferences (e.g., a user likes spicy food) into account. However, to design a "health-aware" system by also focusing on medical aspects, it is of utmost importance to include additional information like nutritional information into the system. To further enhance the system, another goal is to improve the Explainability of recommendations. As explanations can be generated relatively easily via linked data, the focus will be on semantic technologies like RDF **[RDF]**. Unfortunately, there is no standard data set in the food domain as for example in other domains like movie recommendation. Thus, data must be created.

This thesis could have the following coarse "working steps":

1. Data cleaning
    a. Generate data from pre-existing scripts.
    b. Combine the generated data with real-life data (crawled from "food.com")
    c. Transform the resulting data into linked data and combine it with the FoodKG **[Haussmann2019]** (pre-existing linked data)
    d. Simplify the main structures (recipes, user, personal information) by defining an ontology and simplifying/cleaning the data with focus on the performance necessary to domain-typical queries.
    e. Evaluate the performance of the final linked data in comparison to the un-optimized data.
2. Execute recommenders
    a. Find in related work (starting points: **[Yue2021, Kumar2016, Li2020]**) fitting recommenders in terms of relevance and availability of necessary training & testing data
    b. Execute at least 2 recommenders on the data
    c. Evaluate the findings by comparing the results among each other and the results from related work with different data sources

With those steps the following research questions should be answered:

1. How could a general-purpose testing/training linked-data set be created for the domain of food recommendation?
2. How can the data be transformed to optimize the performance when working with recommenders and “explainers”?
3. Which steps must be processed to transform the data “back” into un-linked data?
4. How does the data perform in comparison to related work?
5. How do related recommenders perform on the data in comparison?

The results must be evaluated as described in the “working steps”.

References

[Haussmann2019] Haussmann, S., Seneviratne, O., Chen, Y., Ne’eman, Y., Codella, J., Chen, C. H., … Zaki, M. J. (2019). FoodKG: A Semantics-Driven Knowledge Graph for Food Recommendation. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Vol. 11779 LNCS, pp. 146–162). Springer. https://doi.org/10.1007/978-3-030-30796-7_10

[RDF] Dan Brickley and R.V. Guha. Rdf schema 1.1. Also available at https:// www.w3.org/TR/2014/REC-rdf-schema-20140225/, accessed: 2020-11-20.

[Yue2021] Yue, W., Wang, Z., Zhang, J., & Liu, X. (2021, April 1). An Overview of Recommendation Techniques and Their Applications in Healthcare. IEEE/CAA Journal of Automatica Sinica. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/JAS.2021.1003919

[Kumar2016] Kumar, A., Tanwar, P., & Nigam, S. (2016). Survey and evaluation of food recommendation systems and techniques. In Proceedings of the 10th INDIACom; 2016 3rd International Conference on Computing for Sustainable Global Development, INDIACom 2016 (pp. 3592–3596). Institute of Electrical and Electronics Engineers Inc.

[Li2020] Li, D., & Zaki, M. J. (2020). RECIPTOR: An Effective Pretrained Model for Recipe Representation Learning. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1719–1727). Association for Computing Machinery. https://doi.org/10.1145/3394486.3403223

Supervisors : Felix Bölz, Diana Nurbakova, Sylvie Calabretto

—---

8. Title : Explanation and Persuasion Methods for Food Recommendation

Commenté [1]: Can be split in 2 separate topics

Description :

This topic is located in the domain of food recommender systems. The goal of such a system is to recommend food related items (like recipes) to a given user. Traditional recommender systems in this domain tend just to focus on improving the recommendations. However, in order to work with people in the real world, in addition to those recommendations explanations and persuasion methods must be added to such a system to trigger desired behaviour changes of the user (e.g., eating “healthier”).

The goal of this project is to bridge the gap between a set of recommendations and the targeted user. Thus, the following steps (or similar) should be carried out:

1. Explanation Methods
    a. Related Work (starting points: **[Fu2020, Tintarev2015, Tsai2019]**) should be sighted on different explanation methods, in particular how those “explain” recommendations

and how the explanation is visualized. In this thesis, the question of "how explanations are generated" is out of scope.

   b. To evaluate the most promising methods, a set of recommendations must be artificially generated (or by hand).
   c. The Methods in combination of certain recommendations must be evaluated via an online questionnaire on participants from acquaintances or the university. Thus, especially interactive methods must be implemented to show them on a UI.

2. Persuasion Methods
   a. Related Work (starting points: **[Oinas-Kukkonen2009, Torkamaan2021, Baumeister2019, Pintar2021]**) should be sighted on different persuasion methods for behaviour change in the food recommendation domain.
   b. The most promising approaches must be implemented.
   c. As for the explanation methods, the promising persuasion methods must be evaluated on a live system over a certain amount of time on a small number of participants. In order to process such an evaluation, the following requirements must be complied with:
      i. From Related Works, any working (and personalized) "live" food recommender must be used to generate recommendations.
      ii. The system's outputs (depending on the complexity the thesis might be adapted) should be used to create online UIs for the evaluation for the persuasion methods on each user.

With those steps, the following research questions should be answered:

1. Which explanation methods are the **most promising** ones for a recommender system in the food domain?
2. Are explanations useful at all in this domain based on different user groups (maybe age, gender, medical situation, ...)
3. Which persuasion methods are the **best performing** ones for a recommender system in the food domain?

This thesis includes the necessity to work with applied machine learning and thus prior knowledge will be very helpful. The evaluation should be processed similarly as described above.

References

[Fu2020] Fu, Z., Xian, Y., Gao, R., Zhao, J., Huang, Q., Ge, Y., … De Melo, G. (2020). Fairness-Aware Explainable Recommendation over Knowledge Graphs. In *SIGIR 2020 - Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 69–78). Association for Computing Machinery, Inc. https://doi.org/10.1145/3397271.3401051

[Tintarev2015] Tintarev, N., & Masthoff, J. (2015). Explaining recommendations: Design and evaluation. In *Recommender Systems Handbook, Second Edition* (pp. 353–382). Springer US. https://doi.org/10.1007/978-1-4899-7637-6_1

[Tsai2019] Tsai, C. H., & Brusilovsky, P. (2019). Explaining recommendations in an interactive hybrid social recommender. In *International Conference on Intelligent User Interfaces, Proceedings IUI* (Vol. Part F147615, pp. 391–396). Association for Computing Machinery. https://doi.org/10.1145/3301275.3302318

[Oinas-Kukkonen2009] Oinas-Kukkonen, H., & Harjumaa, M. (2009). Persuasive systems design: Key issues, process model, and system features. *Communications of the Association for Information Systems*, *24*(1), 485–500. https://doi.org/10.17705/1cais.02428

[Torkamaan2021] Helma Torkamaan and Jürgen Ziegler. 2021. Integrating Behavior Change and Persuasive Design Theories into an Example Mobile Health Recommender System. In Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp '21). Association for Computing Machinery, New York, NY, USA, 218–225. https://doi.org/10.1145/3460418.3479330

[Baumeister2019] Baumeister, H., Kraft, R., Baumel, A., Pryss, R., Messner, EM. (2019). Persuasive E-Health Design for Behavior Change. In: Baumeister, H., Montag, C. (eds) Digital Phenotyping and Mobile Sensing. Studies in Neuroscience, Psychology and Behavioral Economics. Springer, Cham. https://doi.org/10.1007/978-3-030-31620-4_17

[Pintar2021] Ana Pintar, Jure Erjavec (2021). A Framework for Designing Behavioural Change with the Use of Persuasive Technology. In International Journal of Management, Knowledge and Learning Vol. 10 (2021). https://doi.org/10.53615/2232-5697.10.75-84

Supervisors : Felix Bölz, Diana Nurbakova, Sylvie Calabretto
—---

9. Title : Press analytics through machine learning for company monitoring

Description :
Contexte scientifique :

It is very important for companies to monitor their own reputation as well as the reputation of their competitors. To automatize this task, machine-learning tools have to be developed. The techniques involved come from Information Retrieval (such as inverted indexes), Machine learning (word embedding through neural networks, nlp techniques, …) and Semantic web technologies. First, a text has to be matched with a company, and then the text can be analyzed to extract structured information. This information is then to be transformed into decisions regarding the stock market for example. We have a large collection f time stamped press articles that can be used as a first dataset. Other data sources can be collected as well by developing

Objectifs :
• to automatically link a press article to a company by developing a score representing the likelihood that a given article is linked to a given company.
• to extract the article subject and classified information about the company, such as: geographic location, key persons, other linked companies.

Méthodologie :
A first step is to dress the state of the art on text mining techniques and then to create a prototype that covers as much as possible the objectives. You will probably use Python as programing language and Elasticsearch as DBMS. Nevertheless, we are open to discuss the concrete subjects within the frame of the topics announced! Depending on the size of the team other topics can be addressed.

Mots-clés
Text mining, Machine learning, NLP, company monitoring
Contexte de travail
This project is in the frame of a collaboration between the LIRIS lab and a company Alteca or Nunki. Alteca side there is at least an engineer (IF 2021 and future PhD) involved. Weekly meetings are to be held, maybe more often at the beginning.

Références

Luan, Y., Ostendorf, M., & Hajishirzi, H. (2018). Scientific Relation Extraction with Selectively Incorporated Concept Embeddings. SemEval Task 7, 7. http://labs.semanticscholar.org/corpus/

Sebastian Ruder. (2019). Relationship Extraction | NLP-progress. http://nlpprogress.com/english/relationship_extraction.html

Wang, H., Qin, K., Zakari, R. Y., Lu, G., & Yin, J. (2021). Deep Neural Network Based Relation Extraction: An Overview. https://arxiv.org/abs/2101.01907v2

SemEval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers, K. Gábor, D. Buscaldi, Anne-Kathrin Schumann, B. Zadeh, Haïfa Zargayouna, Thierry Charnois Computer Science *SEMEVAL 2018, https://arxiv.org/pdf/1808.08643.pdf
Supervisors : Elöd Egyed-Zsigmond, Harald Kosch —---

10. Title : Subjects linked to text mining and data visualization

Description : I have several subjects related to text-mining : event tracking in the online press, argument mining from texts, large document collection visualization. Many of them are in collaboration with companies. Contact me for more details.

Supervisors : Elöd Egyed-Zsigmond