Knowledge-Augmented Induction of Complex Networks on Supply–Demand–Material Data

Dan Hudson[1]⋆, Leonid Schwenke[1]⋆, Stefan Bloemheuvel[2], Arnab Ghosh Chowdhury[1], Nils Schut[3], and Martin Atzmueller[1]

[1] Osnabrück University, SIS Group, Osnabrück, Germany
[2] Tilburg University, Jheronimus Academy of Data Science, 's-Hertogenbosch, The Netherlands
[3] Polymer Science Park, Zwolle, The Netherlands

Abstract. We describe a method for complex network induction in a knowledge-augmented data-driven approach. For this, we match items in a database according to their attributes, using knowledge of sub-contexts within the problem domain to improve the specificity and relevance of matches; this relates specifically to the challenge of supply chain modelling for the recycled plastics industry, using heterogeneous supply-demand-material data. In our approach, knowledge of sub-contexts comes from a mixture of data-driven inference and input from experts, and is crucial in determining how best to match items to one another. We store domain-specific knowledge in the form of patterns that describe subgroups of our data, a 'case base' for use in case retrieval, and also explicit rules provided by experts. We present a system prototype, describe the conceptual modelling approach, and discuss preliminary outputs demonstrating the proposed modelling method. An effective supply chain model can be used to support the recycled plastics industry and expand the uptake of recyclate.

1 Introduction

Supply chains [10, 29] can be defined as all stages involved in producing and delivering a product from supplier to customer – historically considered as a series of steps [16]. However, recent studies have used network theory to model supply chains as complex networks [24, 30]. This requires explicit information on the supply chain elements, which is not always available, a common gap which our system aims to address in order to then model supply chain data as a complex network – in a knowledge-augmented approach.

In the context of the Di-Plast project [19], we focus on utilising industrial *supply–demand–material* data from the recycled plastics industry on suppliers, buyers and products with specific material specifications. However, supplier–product and buyer–product information is only provided in heterogeneous form, which needs to be aligned and matched, leading to *resource-induced* complex networks. This requires a knowledge-augmented network induction approach.

⋆ Both authors contributed equally to this research.

In a data-driven approach, we start with *supplier–product* and *buyer–product* specifications, i. e., complex user–item relationships, where the item part contains complex product and/or material specifications. The central problem we address is the matching [11, 38] of the respective entities, i. e., the complex material specifications of a product, in order to form the complex network/graph structure. This is difficult, due to the complex alignment and the specification of constraints which have to be fulfilled during the matching. However, this can be supported in a data-driven way – mining important parameters – while also including background knowledge of domain experts in order to guide the process. Then, with a human-in-the-loop approach, important (complex) constraints can be captured, domain knowledge e. g., on the importance of features can be provided, and the mined relations and induced networks can be inspected.

Therefore, we propose a knowledge-augmented, context-based and data-driven approach. We use background knowledge, subgroup discovery [4] as well as techniques from case-based reasoning, specifically focusing on case-retrieval. This takes the form of a matching and similarity ranking method for creating the respective edges between buyer and seller nodes in our complex network representation. Essentially, we match them according to their product specifications, utilising the knowledge-augmented matching method. Two special challenges in our industrial context are that many properties can be unknown and often no perfect match exists for a buyer specification and, further, that often the detailed target context is not known. This highlights the need to find the hidden context and use a knowledge augmented similarity approach to find possible alternatives.

Similar products between suppliers and buyers are *matched* and used for *relationship modelling*, leading to our desired complex network abstraction. The complex network itself is facilitated by a tripartite graph representation and/or analysing the respective bipartite projections, as discussed below. For this, we adapt analysis principles of complex networks and folksonomies [20, 15, 40].

Our contributions are summarised as follows:

1. We present a framework for inducing complex networks in order to model heterogeneous supply-demand-material data. The resulting complex networks (or graphs) can then be analysed for a variety of purposes, such as identifying important suppliers in the network, identifying gaps in the supply chain, and developing profiles of the materials required in different industries. For this, we propose a knowledge-augmented data-driven approach for creating the graph structure, i. e., the links between the respective tripartition of supplier–product–buyer nodes.
2. This is supplemented by a first view on our prototypical system implementation in the context of the recycled plastics industry, for which we discuss a preliminary evaluation of our outputs with a domain expert.

The rest of the paper is structured as follows: Section 2 summarises related work. After that, Section 3 describes our proposed approach. Section 4 presents and discusses preliminary results. Finally, Section 5 concludes with a summary and interesting directions for future work.

2   Related Work

Below, we discuss related work on matching and context-aware retrieval: We briefly summarise case-based reasoning (CBR) [35, 1] – an analogical reasoning technique used in various machine learning methods [35, 1]. Next, we introduce subgroup discovery, a technique to find interesting subsets of of data points.

Matching and Context-Aware Retrieval. The problem of matching items is closely related to that of search and recommendation in user–item interactions, where users are interested in obtaining *relevant* items for their queries. This has been approached in numerous ways e. g., [20, 21, 18, 38, 39]. Matches can then be analysed and processed using graph-based [20, 21] approaches, which we focus on, or using deep learning techniques [38]. Furthermore, we focus on different contexts and sub-contexts of the user, which we use for entity matching, as well as relationship modelling in our complex network. This is called context-aware retrieval [13], also relating to context aware queries [8]. In the context of recycled plastics, one big challenge is that producers of certain products want to find alternatives for their non-recycled plastics. Thus, they have a working specification, but no exact recycled alternative exists (buyer context). By considering similar specifications in the same application area from other buyers (hidden context) we want to identify important target attributes for adapting our results.

Case-Based Retrieval. We apply an adapted case-based retrieval [28, 2, 12] method for ranking and retrieving products that match to a particular query, while in our application, the outcomes of each case are not known. [34] implemented a CBR ranking system for products over a user query which can provide suggestions on how the user could change the query to find noteworthy products based on the selections of other users. In our case, the decision process of a buyer is rather complex, taking different constraints into account. Therefore, we need to have an awareness of the buyer's hidden context to understand the applied complex data. How to evaluate such a complex context was researched by [17], on which basis we decided to do an evaluation by experts. A further challenge is the high proportion of missing data. For this reason and based on the research from [26], we try to handle missing values also in a context-based manner.

Subgroup Discovery. Subgroup discovery [4], aims at identifying subgroups of data instances that are *interesting* with respect to a certain target concept, e. g., having a high chance of some interesting attribute being present. A subgroup can be represented by a pattern which specifies rules for membership in the subgroup, typically in the form of feature–value pairs, which must hold true for a data instance in order for it to be included in the subgroup. This means that it is a data-driven process that discovers explicit and *interpretable* rules to associate the target concept to attributes found in the data instances. Subgroup discovery has been applied to, e. g., analyse medical knowledge [32], industrial data [22] and social media [3]. In our work, subgroup discovery is used to identify sub-contexts (hidden contexts) of buyer specifications, where certain attributes may become more important than others in the matching process.

3   Method

Our method targets two objectives: First, to create a ranking of product specification matches, and second, to create a hypergraph using this. To do so, we match buyer property specifications Q (also called *queries*) in the form of attributes to the product attributes A of different sellers. In many cases, no exact match be can be found, however similar products could satisfy the buyer's needs. Normally an expert would analyse the needs and make suggestions based on this. To support an increasing demand for recyclate, an automatic approach is desired. Our method includes a case-based approach to understand which deviations will be acceptable, based on the *queries* from previous buyers who looked for recycled materials in the same application context, i. e., we let the buyer's application context inform the matching process.

Our method is presented diagrammatically in Figure 1 (A). A complex network abstraction is the end goal of our work, which we describe in subsection 3.1. To begin the process, we describe an approach to automatically extracting information to place into the Supplier-Product and Buyer-Product databases, in subsection 3.2. Next, in subsections 3.3 and 3.4, we describe a method for matching product specifications, leading to the relationship modelling that then forms the basis of our final network abstraction. In subsection 3.3, we explain how subgroup discovery can infer a relevant context when matching product specifications, leading to even greater specificity in how different attributes are weighted in the matching process described in subsection 3.4. The remaining steps required to match product specifications and to create the edges for our network model are described in 3.4. This includes a step to transform the feature space according to the context of the buyer, thereby focusing the matching procedure on the most relevant attributes of the products. Concluding with subsection 3.5, we explain how our method supports interpretation, explanation and adaptation of the matching process.
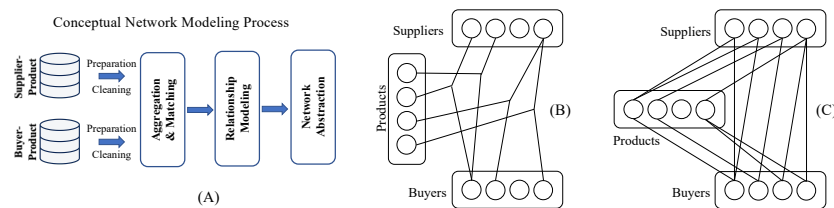
3.1   Complex Network Model

Fig. 1. Conceptual Graph Modeling Process (A), Example Tripartite Hypergraph (B) and projected Hypergraph (C).

To model the supply chain, we consider networks modelled as graphs $G_S, G_B, G,$ with the bipartite supplier and buyer graphs $G_S = (V_S, V_P, E_S), G_B = (V_B, V_P, E_B,$

where V_S and V_B indicate suppliers and buyers and V_P indicates products (with heterogeneous textual/numeric/nominal material specifications from a set $A = \{a_1, a_2, \ldots, a_n\}$, attributed to a node), $E_S \subseteq V_S \times V_P$, $E_B \subseteq V_B \times V_P$, and finally the induced tripartite hypergraph $G = (V_S, V_B, V_P, E)$, where $E \subseteq V_S \times V_B \times V_P$ captures ternary supply–demand relationships between suppliers and buyers for specific products. We thus aim at a tripartite hypergraph between suppliers, products, and buyers. Figure 1 – part (B) – shows an example graph. This graph can then also be projected onto bipartite graphs, as shown in part (C), similar to procedures in folksonomies [20] such that according methods like search, recommendation and ranking can also be directly applied on the graph. This augments the capabilities of our system to adapt based on accumulated knowledge.

3.2   Sources of Data

The data we use to construct our model of the recycled plastics supply chain is taken from our 'Matrix Tool', created in collaboration between Osnabrück University and industrial partner Polymer Science Park (PSP). This contains a database of plastic recyclate suppliers, specifications of the recyclate products they offer, plus potential buyers and the specifications of the products they make from recyclate.

For obtaining supplier–product–buyer data, the Matrix tool relies on data extracted from PDF data sheets containing product specifications. This specifically targets the cold-start problem, i.e., when no (or little) user information is available. On the PDF data sheets we apply document layout analysis, to identify and to analyse the physical and logical document structure to extract relevant
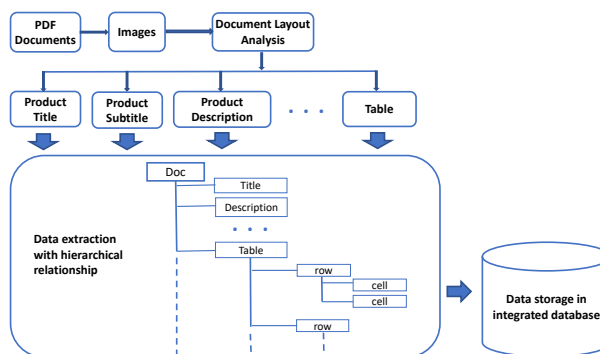

Fig. 2. Overview of data extraction and data integration pipeline (adapted from [33]).

information [27, 37] and also extract information from other heterogeneous data sources. Figure 2 illustrates the extraction process. Supplier-Product and Buyer-Product is stored in an appropriate database, and becomes the basis for inferring Supplier-Product-Buyer edges in our network abstraction.

3.3   Subgroup Discovery

Our method aims to match buyers and products in a context-aware manner. When matching a buyer specification, we use subgroup discovery to find a rele-

vant context, specified through other similar buyer specifications in the database. Subgroup discovery [4], aims at identifying subgroups of data instances (in our case buyer specifications) that are *interesting* with respect to a certain target concept, e. g., regarding a specific processing technology such as injection moulding. Using a binary target concept such as this, we are interested in large subgroups with a high share of instances for which the target concept is true, e. g., being very predictive of a specific production process.

Subgroups are described by a symbolic pattern which is typically given by a conjunction of feature–value pairs in the case of nominal features and selections on intervals in the case of numeric features. In our work, this is performed on the set of product attributes, so that we obtain conjunctions of constraints on the attributes. An example in our application context is given by *'MFI Minimum'* < 1.0 *AND 'Elongation at Break'* = NA *AND 'MFI Maximum* < 1.500. Here, *MFI* indicates a specific property of a material. A pattern can thus also be interpreted as the *body* of a *rule*. The rule *head* then depends on the target concept. In a top-k setting, a subgroup discovery algorithm returns the top-k subgroups according to a selectable interestingness measure, c. f., [4]. For a binary target concept, e. g., the size of a subgroup described by the pattern (its *support*), and the share of the target concept in the subgroup, (its *confidence*), are combined by one of the standard interestingess measures.

Finding subgroups that, for example, have a high probability of being appropriate for a certain manufacturing process, we can identify contexts consisting of highly relevant product specifications. These sub-contexts can then be used to normalise the data in a way that is more relevant to what the buyer is looking for, leading to better matches, as described in subsection 3.4, next.

3.4   Creation of Tripartite Supplier–Product–Buyer Edges

An edge in our tripartite hypergraph model represents a match between a "buyer query" $Q \subseteq A$ of a node V_B specified via a set of attribute constraints (our query on material attributes from the set of parameter attributes A) and the according attributes of a node V_P of a supplier node V_S, as defined above. We aim to build a similarity ranking to indicate matches between queries and product attributes for constructing our graph model. To create edges, we can then take the top-n elements in the ranking, or just the top-1 match for a simple graph.

We suggest a buyer context aware ranking approach to compare multiple related products to one query Q. A context is described by $C \subseteq A \setminus Q$, which consists of several similar buyer specifications. At first an initial limitation of the context is provided from the buyer, but via subgroup discovery we next want to find the most suitable hidden context C_h. An example for a context could be the attribute *product=pipes*, where a more specific unknown hidden context C_h exists e. g., underground pipes, which we need to discover via subgroup discovery. Inside subgroup C_h, different degrees of variance for the individual attributes can be observed. We argue that a smaller variance of an attribute inside a subgroup means that this attribute is especially important for the context C_h. Concluding, an attribute with a lower variance is more important compared to one with

a higher variance. This further highlights the importance to find the correct sub-context via subgroup discovery. Compared to traditional methods, we decide the importance of features according to previous buyer specifications rather than product specifications, and achieve this by discovering a detailed (hidden) context for the matching on a case-by-case basis (see subsection 3.3).

After finding the hidden context, we normalise the data (product, query), by transforming the attribute value space into a Gaussian form based on C_h to achieve a normal distribution and stabilised variances. In this way, we implicitly include the weight/importance of an attribute into the according normalised value space.

This is similar to other work on CBR problems that has used a weighting based on Gaussian distributions [25, 31]. An example of this transformation for the attributes MFI (melting flow index) and density is shown in Figure 3, demonstrating the resulting normalised distribution. We now compute the Euclidean distances between a query and potentially matching products within this transformed space. This normalisation provides a data-driven adaptation of the matching procedure to the context at hand. Hence, an adapted query can now be better assessed inside the normalised feature space. Additionally, our method adapts according to knowledge that has been cap-
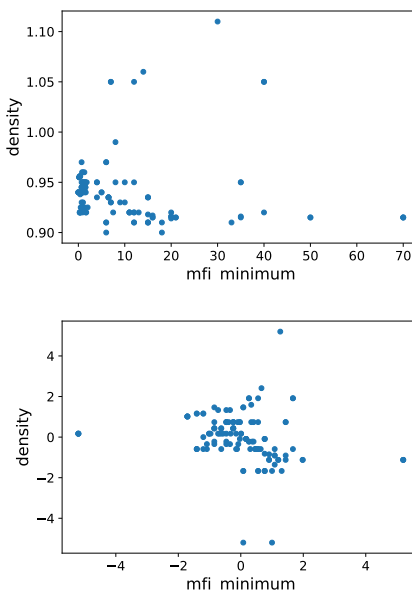

Fig. 3. Example distribution of the attribute *Density* and the attribute *MFI Minimum* before (top) and after (bottom) the Gaussian transformation, showing the shifted distances based on the variance.

tured from experts. To return to the example of *product=pipes*, when constructing piping, there is a need for sufficient rigidity to prevent the pipe from bending. In this instance, expert knowledge may inform the matching procedure by adding the constraint that an e-modulus higher than stated in the query is acceptable while a lower e-modulus is not. We capture this type of knowledge as *complex constraints* for the relevant *context*. Using domain knowledge also makes it possible to define partial matches for non-numeric attributes, e. g., HDPE is a sub-material of PE, which is handled as a match rather than being penalised with an error of 1 in our distance calculations. Alternatively, non-numeric attributes can be included as constraints in the definition of a context C, thus impacting the data subset which normalises the attribute space.

3.5   Interpretation, Explanation, Adaptation

The steps described so far provide a matching between suppliers, products and buyers, that then can provide tripartite edges in a hypergraph. Using this graph structure, as well as the case information contained in the ranking we can provide an explanation on why a product is ranked as it is. We can perform this on different levels: First, we can show related cases. Second, we can visualise the discovered subgroups for a given context C and show how each attribute space is transformed and why. In this space the distance can be visualised and easily interpreted by a human. Further we can give examples based on older searches which illustrate the deviation in the given context to explain the base idea of the weights. Finally, we can also apply techniques of CBR for case-based explanation [14, 23]. In particular, we can also adapt cases utilising the top-n matches of a query utilising the *adaptation step* in the CBR cycle [1], such that we can, e. g., merge cases into prototypical cases for both summarising cases as well as provide explainable candidate cases to the user, c. f., [9, 5].

4   First Results

In this section we present the results from subgroup discovery, evaluate a preliminary example of rankings according to the knowledge from a domain expert and showcase a portion of the induced hypergraph structure.

4.1   Subgroup discovery

As stated in section 3.4, we use subgroup discovery as a way to detect possible unknown sub-contexts within a larger context, such as sub-contexts of the automobile industry. As outlined above, subgroup discovery operates by first specifying a variable of interest (the 'target variable'), and then applying a discovery algorithm that identifies sets of membership criteria, also known as 'patterns', such that the membership criteria identify a collection of instances with some atypical average value, as determined by a quality function. The discovered sub-contexts identify a selection of data points that are closely related and are particularly relevant for the query at hand. Using the VIKAMINE system [6], subgroups were identified within the different markets in the business domain. Below are examples taken from the construction market, where each subgroup is represented by a pattern of membership criteria:

- 'Elongation at Break' = *NA* AND 'OIT' = *NA* AND 'MFI Maximum' < 1.5
- 'MFI Maximum' < 1.5
- 'MFI Minimum' < 1.0 AND 'Elongation at Break' = *NA* AND 'MFI Maximum' < 1.5
- 'MFI Minimum' < 1.0 AND 'Elongation at Break' = *NA* AND 'OIT' = *NA*
- 'MFI Minimum' < 1.0 AND 'MFI Maximum' < 1.5

These patterns indicate that there is a sub-context of the construction market in which elongation at break and OIT (oxidative induction time) are not relevant, and in which MFI values should be low.

Table 1. Example: Top-5 of the ranking for a search query (in bold) in the context of building constructions. For the abbreviation overview see:
`https://www.professionalplastics.com/ACRONYMS?XLT_TO=en`

Similarity	Material	Processing	MFI-Min	MFI-Max	Density	E Modulus	TYS	TYE	OIT
Query:	**HDPE**	**extrusie**	**0.3**	**0.5**	**0.96**	**900**	**25**	**11**	**10**
0.998948	PE	extrusie	1.5	10.0	0.95	850.0	23.0	75.0	NA
0.998948	PE	spuitgieten	1.5	10.0	0.95	850.0	23.0	75.0	NA
0.997435	HDPE	extrusie	0.7	0.7	0.97	560.0	24.0	NA	NA
0.996710	HDPE	extrusie	1.3	1.3	0.96	950.0	28.0	NA	NA
0.996667	HDPE	extrusie	1.5	1.5	0.95	900.0	25.0	NA	NA

4.2 Example: Ranking-Induced Graph

Subgroup discovery helped us to find smaller clusters in our given context. For enabling a knowledge-augmented approach, we also included some initial expert knowledge for the selection of important features. In further steps and with more data we want to develop a semi-automatic selection of the important product features.

Table 1 depicts an example ranking for a given query. The rows indicate different product specifications. According to inspection by a domain expert, the produced top rows are relevant, in particular rows 1-5 all include relevant specifications. However, rows 3-5 are slightly more relevant than the others, since for some parameters, the deviations between query and provided values should only deviate in one direction. So, ultimately the ranking needs to be reordered using some additional domain constraints.

Fig. 4. Tripartite HyperGraph example from our dataset where blue nodes are materials (m), green are buyers (b) and red are suppliers (s).

This is an example of the domain knowledge that needs to be incorporated into our knowledge augmented approach. It is important to note, however, that the domain knowledge required to match products to buyers is quite complex, and further work is needed to capture and exploit this knowledge. In our approach, we can either incorporate this using domain knowledge, or by enriching the graph using multi-edges for capturing a larger set of matched options. Finally, Figure 4 shows an example visualisation of the hypergraph created from a subset of the real-world data.
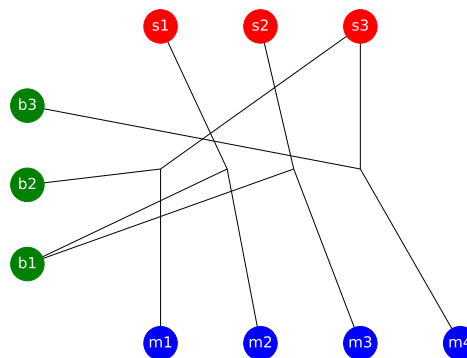
5   Conclusions and Future Work

In this work, we focused on utilising complex industrial *supply–demand–material* data for context-based search and ranking, which we also implemented in a system prototype. We presented a framework for knowledge-augmented induction of complex networks for modeling complex relationships in the context of heterogeneous data. Our hypergraph modelling approach can be generally applied on supply chains with Supplier-Product-Buyer relations. Our proposed matching process is suited to domains where previous buyer requests are informative about (hidden) buyer contexts which in turn are informative about the importance and availability of attributes used for matching. In the application domain of recycled plastics, our proposed knowledge-augmented data driven approach showed first promising results according to the assessment of domain experts.

Future steps include further domain specific fine-tuning of the matching, incorporating data about the selection step of real users to enable a fine-grained application of CBR approaches, in particular taking the adaptation step of CBR into account. This also concerns the further formalisation and inclusion of domain knowledge into the proposed framework, in order to enable a refined human-centred knowledge-based approach using specific constraints of real users. In addition, we intend to investigate further refinements of the hypergraph model, as well as augment the hypergraph model further, leading to knowledge graph structures, such that both knowledge modeling as well as application can be integrated into the same structural representation, e. g., [7, 36].

Last but not least, we aim to analyse the industrial data with multiple complex network methods. For example, (1) link prediction could be performed on the supply-demand-material graph to infer new edges, and (2) community detection can help with identifying new subgroups in the data. In addition, (3) global graph metrics such as density and the average degree of the nodes in the hypergraph could provide further insights into the characteristics of the data.

Acknowledgements

References

1. Aamodt, A., Plaza, E.: Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AI Communications **7(1)**, 39 – 59 (1994)
2. Arnold, C.W., El-Saden, S.M., Bui, A.A., Taira, R.: Clinical case-based retrieval using latent topic analysis. In: AMIA annual symposium proceedings. vol. 2010, p. 26. American Medical Informatics Association (2010)
3. Atzmueller, M.: Mining social media: key players, sentiments, and communities. Wiley Interdisciplinary Reviews: DMKD **2**(5), 411–419 (2012)
4. Atzmueller, M.: Subgroup Discovery. WIREs Data Mining and Knowledge Discovery **5**(1), 35–49 (2015). https://doi.org/10.1002/widm.1144

5. Atzmueller, M., Baumeister, J., Puppe, F.: Evaluation of two Strategies for Case-Based Diagnosis handling Multiple Faults. In: Proc. 2nd Conf. Professional Knowledge Management (WM2003). Luzern, Switzerland (2003)
6. Atzmueller, M., Lemmerich, F.: VIKAMINE - Open-Source Subgroup Discovery, Pattern Mining, and Analytics. In: Proc. ECML/PKDD. Springer, Berlin, Germany (2012)
7. Atzmueller, M., Sternberg, E.: Mixed-Initiative Feature Engineering Using Knowledge Graphs. In: Proc. 9th International Conference on Knowledge Capture (K-CAP). ACM Press, New York, NY, USA (2017)
8. Bai, J., Nie, J.Y., Cao, G., Bouchard, H.: Using query contexts in information retrieval. In: Proc. annual international ACM SIGIR conference on Research and development in information retrieval. pp. 15–22. ACM, New York, NY, USA (2007)
9. Baumeister, J., Atzmueller, M., Puppe, F.: Inductive Learning for Case-Based Diagnosis with Multiple Faults. In: Advances in Case-Based Reasoning. LNAI, vol. 2416, pp. 28–42. Springer, Berlin, Germany (2002)
10. Beylot, A., Villeneuve, J.: Assessing the national economic importance of metals: An input–output approach to the case of copper in france. Resources Policy **44**, 161–165 (2015)
11. Blanco, R., Cambazoglu, B.B., Mika, P., Torzec, N.: Entity recommendations in web search. In: International Semantic Web Conference. pp. 33–48. Springer (2013)
12. Brown, M.G.: An underlying memory model to support case retrieval. In: Proc. EWCBR. pp. 132–143. Springer, Heidelberg, Germany (1993)
13. Brown, P.J., Jones, G.J.: Context-aware retrieval: Exploring a new environment for information retrieval and information filtering. Personal and Ubiquitous Computing **5**(4), 253–263 (2001)
14. Caro-Martinez, M., Recio-Garcia, J.A., Jimenez-Diaz, G.: An algorithm independent case-based explanation approach for recommender systems using interaction graphs. In: Proc. ICCBR. pp. 17–32. Springer (2019)
15. Cattuto, C., Schmitz, C., Baldassarri, A., Servedio, V.D., Loreto, V., Hotho, A., Grahl, M., Stumme, G.: Network Properties of Folksonomies. AI Communications **20**(4), 245–262 (2007)
16. Crandall, R.E., Crandall, W.R., Chen, C.C.: Principles of supply chain management. CRC Press (2014)
17. Gu, M., Aamodt, A.: Evaluating cbr systems using different data sources: A case study. In: Proc. ECCBR. pp. 121–135. Springer (2006)
18. Hinz, O., Eckert, J.: The impact of search and recommendation systems on sales in electronic commerce. Business & Information Systems Engineering **2**(2), 67–77 (2010)
19. van den Hoogen, J., Bloemheuvel, S., Atzmueller, M.: The Di-Plast Data Science Toolkit – Enabling a Smart Data-Driven Digital Circular Economy for the Plastics Industry. In: Proc. DBDBD. JADS, 's-Hertogenbosch, The Netherlands (2019)
20. Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Information Retrieval in Folksonomies: Search and Ranking. In: Proc. ESWC. pp. 411–426. Springer, Heidelberg, Germany (2006)
21. Jäschke, R., Marinho, L., Hotho, A., Schmidt-Thieme, L., Stumme, G.: Tag recommendations in folksonomies. In: Proc. ECML/PKDD. pp. 506–514. Springer, Heidelberg, Germany (2007)
22. Jin, N., Flach, P., Wilcox, T., Sellman, R., Thumim, J., Knobbe, A.: Subgroup discovery in smart electricity meter data. IEEE Transactions on Industrial Informatics **10**(2), 1327–1336 (2014)

23. Jorro-Aragoneses, J.L., Caro-Martínez, M., Díaz-Agudo, B., Recio-García, J.A.: A user-centric evaluation to generate case-based explanations using formal concept analysis. In: Proc. ICCBR. pp. 195–210. Springer (2020)
24. Kim, Y., Choi, T.Y., Yan, T., Dooley, K.: Structural investigation of supply networks: A social network analysis approach. Journal of Operations Management **29**(3), 194–211 (2011)
25. Li, H., Sun, J.: Gaussian case-based reasoning for business failure prediction with empirical data in china. Information Sciences **179**(1-2), 89–108 (2009)
26. Löw, N., Hesser, J., Blessing, M.: Multiple retrieval case-based reasoning for incomplete datasets. Journal of biomedical informatics **92**, 103127 (2019)
27. Marinai, S.: Learning algorithms for document layout analysis. In: Handbook of Statistics, vol. 31, pp. 400–419. Elsevier (2013)
28. Montani, S., Portinale, L., Leonardi, G., Bellazzi, R., Bellazzi, R.: Case-based retrieval to support the treatment of end stage renal failure patients. Artificial Intelligence in Medicine **37**(1), 31–42 (2006)
29. Moran, D., McBain, D., Kanemoto, K., Lenzen, M., Geschke, A.: Global supply chains of coltan: a hybrid life cycle assessment study using a social indicator. Journal of Industrial Ecology **19**(3), 357–365 (2015)
30. Nuss, P., Chen, W.Q., Ohno, H., Graedel, T.: Structural investigation of aluminum in the us economy using network analysis. Environmental science & technology **50**(7), 4091–4101 (2016)
31. Park, Y.J., Kim, B.C., Chun, S.H.: New knowledge extraction technique using probability for case-based reasoning: application to medical diagnosis. Expert systems **23**(1), 2–20 (2006)
32. Puppe, F., Atzmueller, M., Buscher, G., Huettig, M., Lührs, H., Buscher, H.P.: Application and evaluation of a medical knowledge system in sonography (sonoconsult). In: Proc. ECAI. pp. 683–687 (2008)
33. Rausch, J., Martinez, O., Bissig, F., Zhang, C., Feuerriegel, S.: Docparser: Hierarchical document structure parsing from renderings. In: 35th AAAI Conference on Artificial Intelligence (AAAI-21)(virtual) (2021)
34. Ricci, F., Venturini, A., Cavada, D., Mirzadeh, N., Blaas, D., Nones, M.: Product recommendation with interactive query management and twofold similarity. In: Proc. ICCBR. pp. 479–493. Springer, Heidelberg, Germany (2003)
35. Sànchez-Marrè, M.: Principles of case-based reasoning
36. Sternberg, E., Atzmueller, M.: Knowledge-Based Mining of Exceptional Patterns in Logistics Data: Approaches and Experiences in an Industry 4.0 Context. In: Proc. ISMIS. LNCS, Springer, Berlin, Germany ((accepted) 2018)
37. Subramani, N., Matton, A., Greaves, M., Lam, A.: A survey of deep learning approaches for ocr and document understanding. arXiv preprint arXiv:2011.13534 (2020)
38. Xu, J., He, X., Li, H.: Deep learning for matching in search and recommendation. In: The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval. pp. 1365–1368 (2018)
39. Zamani, H., Croft, W.B.: Learning a joint search and recommendation model from user-item interactions. In: Proc. WSDM '20. pp. 717–725 (2020)
40. Zhao, J., Zhang, Q., Sun, Q., Huo, H., Xiao, Y., Gong, M.: Folkrank++: An optimization of folkrank tag recommendation algorithm integrating user and item information. KSII Transactions on Internet and Information Systems **15**(1), 1–19 (2021)