Douglas Cunningham, Petra Hofstedt, Klaus Meer, Ingo Schmitt (Hrsg.): INFORMATIK 2015
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2015

TrendFashion -A Framework for the Identification of Fashion Trends

Samaneh Beheshti-Kashi[1], Michael Lütjen[2], Lennard Stoever[3], Klaus-Dieter Thoben[4]

Abstract: The fashion industry faces different challenges regarding accurate forecasts for future fashion products. The consumer demand is volatile and sales periods of fashion products are short due to production plants in Asia and target markets in Europe. Besides standard statistical approaches based on historical data and advanced methods such as the application of artificial neural networks or fuzzy logic, there are fashion experts, who use different information sources, e.g. fairs, social media, fashion websites, to predict design-trends as well as sales volumes. In this paper we follow this expert-driven approach by collecting data from fashion weblogs, news sites and fashion magazines, in order to identify actual and future design- trends. For this aim, we develop the TrendFashion Tool which collects data from these fashion sources and analyse them. On a higher level, this tool successfully separates fashion related posts from non-fashion related posts. And on a lower level, it identifies fashion related words and weights them according to an index.

Keywords: Fashion Forecasting, Social Media Analytics, Trend identification

1 Challenges in the Fashion Industry

The fashion industry has to handle diverse idiosyncrasies on different levels in order to avoid overstocked and stock-out inventories. External factors such as changing weather conditions, holidays or (sports) events have an impact on short-term customers' purchasing decisions and buying behaviours [Th10]. The availability of a product is crucial for fashion items, since buying decisions are often made at the point of sale [NGP13]. Moreover, different design trends are presented in fashion shows with a long-term range of approximately six months. Besides, street trends are emerging spontaneously with a mid-term range of approximately three months. Additionally, the most fashion products are shorter lived and will stay only for some weeks at the stores.

1 Universität Bremen, International Graduate School for Dynamics in Logistics (IGS), Hochschulring 20, 28359 Bremen bek@biba.uni-bremen.de

2 BIBA - Bremer Institut für Produktion und Logistik GmbH at the University of Bremen, Hochschulring 20, 28359 Bremen, Germany, ltj@biba.uni-bremen.de

3 minubo GmbH, Deichstraße 48–50, 20459 Hamburg, lennard@minubo.com

4 Faculty of Production Engineering, University of Bremen, Germany, tho@biba.uni-bremen.de

Product related factors such as the high variety in sizes and colours intensifies the demand uncertainty [CLP04]. In this regard, we can distinguish between traditional and fast fashion retailing. While traditional fashion retailing consists of producing one collection for the spring/summer and one for the autumn/winter season by producing the products in Asia and shipping them to Europe, the Spanish retailer Inditex is an excellent example of fast fashion retailing. Zara, the major division of Inditex has become the leader in rapid development of fast changing fashions by addressing latest street trends with products made in Spain with lead times less than several weeks. On the other hand, most production plants are placed in Asian countries where products can be produced cheaply [MTD11], but the time-to-market is long. Therefore, the reproduction of good selling products is rarely possible [Fs01]. One option for companies is to fly their goods to Europe or install additional production plants in Turkey or North Africa in order to be time efficient. However, the traditional as well as the fast fashion retailing strategy face challenges, which are worthy considering new approaches of trend identification and forecasting.

2 Related Work

2.1 Fashion Forecasting

The literature describes different approaches within the field of fashion forecasting. Due to the described characteristics of the fashion industry and the lack of historical data, standard forecasting methods such as simple statistical approaches often fail in providing accurate forecasting results. In recent approaches more advanced methods such as fuzzy logic or artificial neural networks are considered [SAC07] [Th10]. Further learning algorithm such as the extreme learning machine [Sz08] or the evolutionary neural network [ACY08] obtains accurate results. Though, in particular hybrid models achieve reasonable accuracy [WG10] [Ct14] [Tm14]. However, these works do not consider further influencing factors in their models. In contrast, Thomassey [Th10] provided a complex approach including external information such as weather conditions or information on promotions and events [Th10]. Considering these additional information, the accuracy of forecast results will increase, since the consumer demand is impacted by exactly these diverse factors. Another approach is followed by Mostard et al. [MTD11] and Teucke et al. [Tm14]. Both focus on pre-order information and include them into their forecasting models. Within the fashion industry the selection of the right colour is a highly important decision [Ki12]. Accordingly, this problem is also addressed by researchers. [Ct12] introduced a model for colour forecasting with very little data. In Gu et al. [GL10] a computer assisted colour forecasting database is presented. Beheshti-Kashi et al. [BT14] suggest integrating social media applications such as Twitter or weblogs into the fashion forecasting process [BT14]. This paper follows this approach and suggestion and introduces the so-called TrendFashion Framework.

2.2 Data Driven Approaches in the Fashion Industry

Within the last 5 years the fashion world has undergone a huge change. In particular, within the last 2-3 years new concepts such as outfittery[5], edited[6], AboutYou[7] or Zalon[8], Zalando's curated shopping solution, have emerged. All these applications require more consumer engagement, for instance, providing more personal data than for a traditional online purchase. Consequently, the companies obtain far more information about their customers than before. Information regarding each purchase, on colours, sizes, number of purchases, preferences, which products have been purchased together, or just product views are easily accessible. Consequently, a huge number of data is available about the consumer for the companies. The availability of these data, paves the way for data driven approaches within the fashion industry. In addition, to the available information provided by the consumer, further data is available throughout the internet. In recent times some start-ups[9] [10] have emerged who focus on these kinds of fashion related data.

Other approaches focus on a typical problem occurring in the fashion industry: sizes and the perfect fit. Every brand has its own particular sizes. Therefore, each garment or shoe ideally should be fitted before purchasing, which is not always possible, in particular in the case of online shopping. Recently, applications such as virtual try-on technologies or 3D Scanners have emerged which can be used by the consumers also at home in order to build their digital avatar. Both technologies are used for ensuring a more accurate fit for the consumer. In addition, to these applications augmented reality environment have obtain increased relevance. These environments allow the user for instance to browse a catalogue and try on apparels [BA15]. Ashdown and Loker consider the virtual try-on as a promising 3D technology for enhancing mass customization and online sales [AL10].

Google as the market leader in search engine technologies has conducted a study on fashion trends considering search queries. In the "Google Fashion Trends Report (U.S.)" they publish their results examining six billion apparel related queries in the United States. They categorize the extracted trends based on user search behaviour into the following six clusters applying Time Series Clustering: Sustained Growth, Seasonal Growth, Rising Stars, Sustained Decline, Seasonal Decline and Falling Stars [ZH15]. This work only considers search queries and not free and unstructured text. Therefore, no automated text processing and Natural Language Processing (NLP) methods are applied. Whereas, considering weblogs and digital fashion magazines require conducting NLP methods, in order to obtain valuable information.

[5] http://www.outfittery.de/
[6] http://www.edited.de/
[7] http://www.aboutyou.de/
[8] https://www.zalon.de/
[9] http://www.fashiongps.com/
[10] https://poshly.com/

2.3 Text Mining and Natural Language Processing

Text Mining methods and approaches have gained increased relevance within recent years. One reason is the huge emergence of free and unstructured data on the web. In particular, due to the numerous social media applications, which empowered every individual user to publish content, Text Mining is required in order to process the data and to extract valuable information. However, Text Mining methods are not limited to web usages, any textual database can profit from Text Mining. Text Mining includes several fields such as information retrieval, text analysis, information extraction, clustering, categorization, visualization, database technology, machine learning and data mining [TA99].

Granitzer [GR06] has identified the following four processes in Text Mining:

- Pre-processing: lexical analysis, part-of-speech tagging, stemming, stop word removal
- Information extraction: machine learning and linguistic analysis
- Feature generation: statistical analysis such as frequency analysis, word co-occurrences
- Operations on feature spaces: clustering, classification etc.

Within the last years the focus of Text Mining has gone towards Natural Language Processing (NLP). NLP is an older research field and involves technologies such as word stemming, lemmatization, multiword phrase grouping, synonym normalization and part-of-speech tagging [KP04].

3 The TrendFashion Framework

TrendFashion is a tool for automated collection and processing of fashion related web data. The current implementation focuses on German language weblogs, fashion magazines and news sites. Following the approach of identifying new design and street fashion trends by analysing web data, such as social media, news sites or digital fashion magazines different challenges have to be managed before the actual integration takes place. One essential task is to monitor the relevant sources. For instance, by continuously collecting and storing data from weblogs, we obtain an image of the fashion blogosphere within a certain time period. However, this data often includes unstructured text data and has to be pre-processed in order to obtain valuable output such as certain colours or fashion styles from it. Figure 3.1 illustrates the concept behind the tool. Within the database we have the following tables: websites, posts, single words and relation as well as the thesaurus. The websites table includes a list of URLs from weblogs, fashion magazines and news sites. With the help of RSS feeds and the crawling functionality we collect the data from these external sources and store them into the posts table. However, not all of the information is stored; we only consider the title, the published date, the URL and the actual content which has usually the format of free text (unstructured text).

Within the data processing phase, the content is analysed by the mean of NLP methods, and the paragraphs are separated into sentences. At this stage, we have included a "Moderelevanz" Index which weights the fashion relevancy of words and sentences. Relevant sentences are those which include for instance colours, brands or textile names. These sentences are separated into words. These words are matched with the thesaurus and a "Moderelevanz" Index is assigned to them. In the following sections, the collecting, processing as well as the thesaurus functionalities are presented.

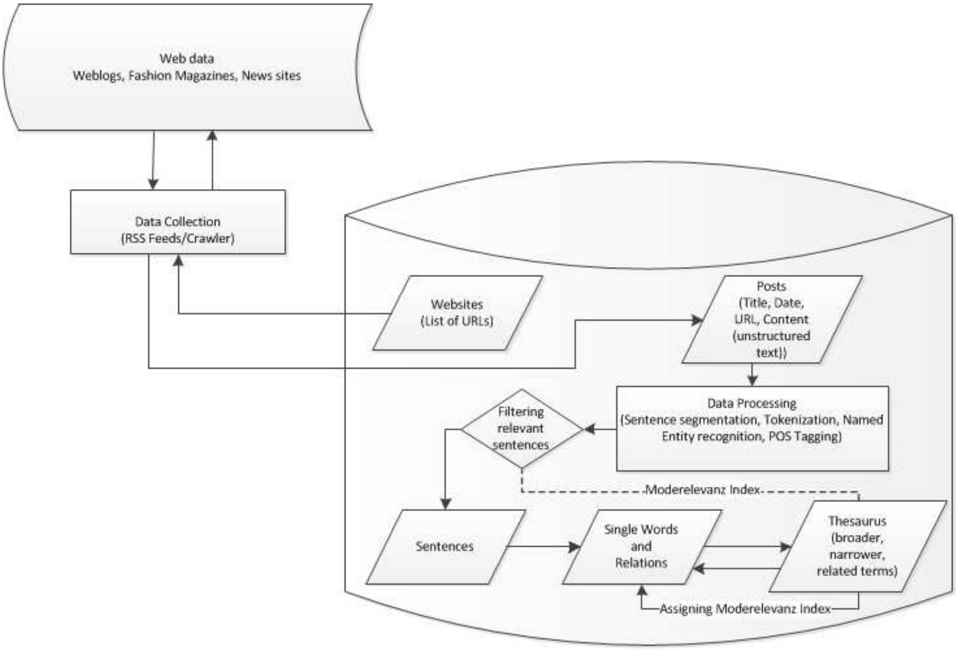

Fig. 3.1: Concept of TrendFashion

3.1 Collecting the Data

The tool is designed to handle different formats, namely RSS feeds and HTML-websites. The first approach was to use RSS feeds exclusively, since they always provide a complete time stamp. A detailed time reference to the written text is highly relevant for the given objective. The significance of the time aspect is due to the fast paced and rapidly changing fashion trends. Therefore, we have to focus on data from the same time period in order to correctly aggregate the data and to conclude the right information. Since a large number of the blogs do not provide RSS feeds, we added the second approach, HTML formats, to the tool. Currently, the tool considers 100 German language weblogs, fashion magazines and news sites.

3.2 Processing the Data

After collection, the data has to be processed. For this task, we apply NLP methods such as sentence segmentation, tokenization, named entity recognition and Part-of-Speech-Tagging. The basis for further processing is the single posts which are stored as single documents. In the first step, the posts are separated into single sentences looking based punctuations. Then, the tokenization process will follow. After separating the sentences, the syntactic analysis starts. The Part-of-Speech-Tagger (POS-Tagger) annotates of every single word of sentence to a specific part-of-speech such as noun, verb or adjective. For the Tagging process we selected the German version of Wiktionary.org, which is a free online wiki-based dictionary, for identifying the correct part-of-speech. The “Moderelevanz” Index, filters and weight the separated POS according to their relevance to fashion topics and filter out stop words such as articles or pronouns at the same time. The extracted and tagged words are stored in a separated table.

3.3 Thesaurus

In order to enhance the semantic functionality of the tool, we have included a thesaurus. Such a thesaurus is often designed for a particular domain, and in contrast to a dictionary it provides relations between words. These relations are required for automated text analysis and are based on several taxonomies in which we classified different product groups. So far, we have included taxonomies only on a high level. Detailed taxonomies are in progress and will be integrated for enhancing the learning capability of the tool. The thesaurus includes broader, narrower and related terms. In addition, to these corresponding terms, the words are stored according to their ”Moderelevanz” Index. Furthermore, the part of speech of each entry is provided. In the current version of the tool, the “Moderelevanz” Index is adjusted and adapted manually.

4 Results

The TrendFashion Database handles different tasks. It collects and processes the data. It filters the content from the noise of a website (navigation, HTML tags etc.) and stores each post separately. Based on the posts, sentences are separated; from these sentences words and word combinations are extracted and stored. Besides the collection and storage of posts, sentences, and words, it provides lists of word occurrences. This is the first step in order to extract future trends. Though, a further focus is on extracting words belonging together, in order to increase the semantic ability of the tool. For identifying fashion trends, adjectives such as colours are crucial. Therefore, it is important to find adjective-noun or adjective-adjective-noun combinations belonging together. In the current version, the tool is able to identify and extract such combinations in a reasonable way. Figure 4.1 illustrates matched word combinations in the area of evening robe, glitter and sparkling clothing. The font size of the word combinations correspondents to their frequency rates: the larger the font size, the higher is their occurrences within the text. In the most cases, the tool matched adjective-noun combinations. However, in the example of *femininenSmokingLook,* we can see that in a first step the two nouns *Smoking* and *Look* are combined, and in the second step the adjective *feminine* is matched to the new combination.

Fig.4.1 Extracted word combinations

With the word cloud in Figure 4.1 a broader and more general view on the extracted word combinations is presented. However, the tool can also identify more specific combinations and relations. Since in fashion just a slight variation of a certain colour, shape or cut will have an impact on the success of a product, we assume that for the prediction of a future design trend more specific features on the colour, cut or style are required and not just a general overview of a product. For instance, looking at the product "blouse", the following adjective noun combinations are extracted *bodenlange Bluse*, *rosafarbene Bluse*, *bieder Bluse, hochgeschlossene Bluse.* In this example, different features describing a "blouse" are provided: a length (*bodenlange*), a colour (*rosafarbene*), a shape (*hochgeschlossene*) as well as information on the style (*bieder*). Similarly, the tool extracts diverse adjective noun combinations for the product jeans. We consider the presented tool as the basis for future predictions. Since the objective is to provide trend information also on a detail level, it is crucial to classify the extracted information according to the different features. This will be also done by the help of taxonomies. Figure 4.2 illustrates the extracted word combinations as a hierarchal tree. In this case, we have taken the features colour, cut, style as well as additional describing attributes and matched the corresponding adjectives to them.

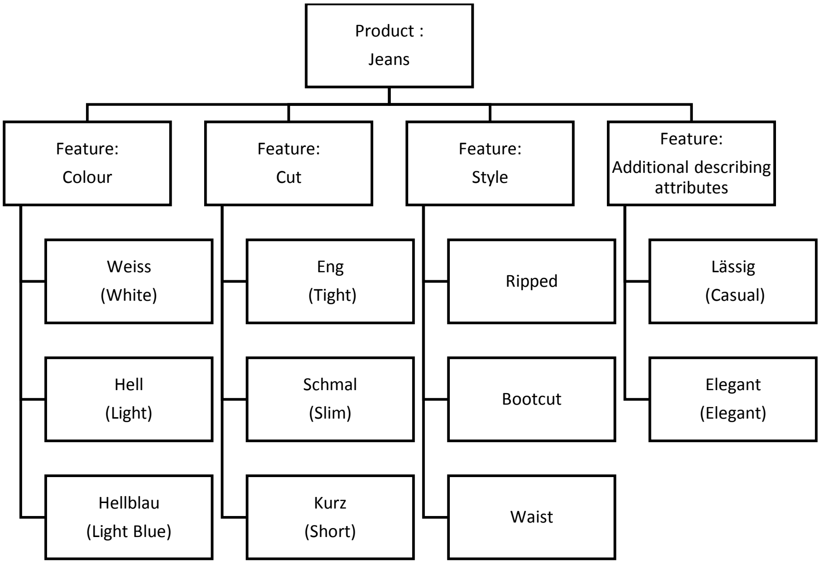

Fig. 4.2: Extracted features on the product jeans

Furthermore, we examined 22 webpages about information on future design trends for the autumn winter season 2015/2016. Figure 4.3 illustrates the identified word combinations extracted from the text provided on the 22 webpages. In contrast to figure 4.2, it shows that more detailed information on a certain product can be extracted. We can find information on the material (*FellPelz*), on certain components of a shoe such as the sole (*dickeSohlen*) or on the heel (*hochAbsatz* or *flacheSchuh*).

Fig.: 4.3: Extracted word combinations on shoe features

5 Conclusion and Future Work

In this paper we propose a data driven approach to predict future design trends. The presented tool handles collecting and processing free and unstructured text data from the web. For this paper, we consider German language weblogs, news sites and digital fashion magazines. The data is collected, stored and analysed by the mean of NLP methods. Extracting and matching of adjective noun combinations are conducted automatically. With the included “Moderelevanz” Index fashion related content can be filtered from the noise to a certain level and is currently manually adapted. The automatic assignment of the Index is considered as the next step. Furthermore, we fill the thesaurus with more detailed taxonomies in order to enhance the semantic functionality. In addition, the time dimension will be added to the analysis in order to be able to build up the trends and to follow them. In the next step, the number of input websites will be increased, in order to work on a larger sample and consequently covering potential niches. Further future work additionally includes extending to English web data, and also to focus on a multi-language approach.

References

[ACY08] Au, K.-F., Choi, T.-M., & Yu, Y. : Fashion retail forecasting by evolutionary neural networks. International Journal of Production Economics, 114(2), 615–630, 2008.

[AL10] Ashdown, S., Loker, S.; Mass-Customized Target Market Sizing: Extending the Sizing Paradigm for Improved Apparel Fit. Fashion Practise 2/2, 147-174.

[BA15] Baytar, F., Ashdown, S.: An Exploratory Study of Interaction Patterns around the Use of Virtual Apparel Design and Try-on Technology. Fashion Practise: The Journal of Design, Creative Process and the Fashion Industry, 7/1, 31-52, 2015.

[BT14] Beheshti-Kashi, S., Thoben, K.-D.: The usage of social media text data for the demand forecasting in the fashion industry fashion sales forecasting and the predictive power of online. In Dynamics in logistics. Fourth international conference, LDIC 2014 Bremen, Germany, February 2014 Proceedings, 1–4, 2014.

[CLP04] Christopher, M., Lowson, R., Peck, H.: Creating agile supply chains in the fashion industry. International Journal of Retail & Distribution Management, 32/8, 367–376, 2004.

[Ct14] Choi, T., Hui, C., Liu, N., Ng, S., & Yu, Y.: Fast fashion sales forecasting with limited data and time. Decision Support Systems, 59, 84–92, 2014.

[Ct12] Choi, T., Hui, C., Ng, S., & Yu, Y.: Color Trend Forecasting of Fashionable Products with Very Few Historical Data, 42/6, 1003–1010, 2012.

[Fs01] Fissahn, C.: Marktorientierte Beschaffung in der Bekleidungsindustrie - Schriften zur Textilwirtschaft, Band 53,. FATM, Münster, 2001.

[Gr 06] Granitzer, M.: KnowMiner: Konzeption und Entwicklung eines generischen Wissenserschliessungsframeworks, PhD Thesis TU Graz, Graz, 2006.

[GL10] Gu, W., & Liu, X.: Computer-Assisted Color Database For Trend Forecasting, 0–3, 2010.

[Ki12] King, J. A.: Colour trends and selection in fashion design. Colour design: Theories and applications, 535–550, Woodhead Publishing Limited, 2012.

[KP04] Kao, A., Poteet, S.: Text Mining and Natural Language Processing – Introduction for the Special Issue. SIGKDD 7/1, 2004.

[MTD11] Mostard, J., Teunter, R., & de Koster, R.: Forecasting demand for single-period products: A case study in the apparel industry. European Journal of Operational Research, 211/1, 139–147, 2011.

[NGP13] Nenni, E. M., Giustiniano, L., & Pirolo, L..: Demand Forecasting in the Fashion Industry: A Review. International Journal of Engineering Business Management, 5, 2013.

[SAC07] Sun, Z.-L., Au, K.-F., & Choi, T.-M. : A neuro-fuzzy inference system through integration of fuzzy logic and extreme learning machines. In: IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics : A Publication of the IEEE Systems, Man, and Cybernetics Society, 37/5, 1321–31, 2007.

[Sz08] Sun, Z.-L., Choi, T.-M., Au, K.-F., & Yu, Y.: Sales forecasting using extreme learning machine with applications in fashion retailing. Decision Support Systems, 46/1, 411–419, 2008.

[TA99] Tan, A.-H.:Text Mining: The state of the art and the challenges, PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases, Beijing, (1999), 65-70, 1999.

[Th10] Thomassey, S.: Sales forecasts in clothing industry: The key success factor of the supply chain management. International Journal of Production Economics, 128/2, 470–483, 2010.

[Th14] Thomassey, S.: Sales Forecasting in Apparel and Fashion Industry: A Review. In. (Choi, T.M. et al.): Intelligent Fashion Forecasting Systems: Models and Applications. Springer-Verlag Berlin Heidelberg 2014, 9-27, 2014.

[Tm14] Teucke, M., Ait-alla, A., El-berishy, N., Beheshti-Kashi, S., & Lütjen, M.: . Forecasting of Seasonal Apparel Products. In (H. Kotzab, J. Pannek, & K.-D. Thoben): Dynamics in Logistics. Fourth International Conference, LDIC 2014 Bremen, Germany, February 2014 Proceedings. Springer, 2014.

[WG10] Wong, W. K., & Guo, Z. X.: A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. International Journal of Production Economics, 128/2, 614–624, 2010.

[ZH15] Zimmer, O., Horwitz, Y.: Google Fashion Trends Report (U.S.), 2015.