User-sentiment based Evaluation for Market Fitness Trackers
Evaluation of Fitbit One, Jawbone Up and Nike+ Fuelband based on Amazon.com Customer Reviews

Hassan Issa, Alaa Shafaee, Stefan Agne, Stephan Baumann and Andreas Dengel

German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany

Keywords: Activity Monitoring, Fitness Applications, Life Logging, Quantified Self, Sentiment Analysis, Fitness Trackers Evaluation, Fitness, Fitbit One, Jawbone Up, Nike+ Fuelband.

Abstract: Wearable fitness and health trackers have been in an accelerated growth since their recent introduction to consumer markets. Given the growth potential of this market sector, much more devices, with unbounded set of features, are being introduced to the consumer at a fast pace. This makes the task of evaluating fitness trackers extremely challenging knowing that the results of an evaluation will quickly become obsolete. In this paper, a user-sentiment based evaluation of fitness trackers is demonstrated on market leading fitness trackers. The used approach relies on the crowd, expressed by Amazon.com product reviews, to present an aspect-based evaluation of any market fitness tracker. Utilizing the crowd knowledge acquired, a personalized recommender system for fitness trackers is also presented.

1 INTRODUCTION

Ever since the recent introduction of wearable computing devices, the market size for these devices has been rapidly growing and expected to attain a market size of $30 billion as forecasted by (BCCResearch, 2014). Accordingly, fitness and health trackers, which correspond to a big part of this growth, are expecting a 96 million units shipment in 2018. Fitness trackers are always-on devices that provide basic pedometer functionalities including estimating steps taken, distance traveled and calories burned. In addition, these devices connect and sync their data easily to smartphone apps and web portals and enable social sharing and friendly competitions. Fitness and health trackers are constantly incorporating new sensors and algorithms to provide further health and fitness insights and functionalities for users. Samsung Simband 2[1] notably collects real-time biometric data including heart rate, blood flow and pressure, skin temperature, CO_2 and oxygen levels as well as EKG[2] levels. Likewise, machine learning techniques are employed by some trackers like AMIIGO[3] fitness tracker which is able to detect over 100 exercises along with the number of sets and repetitions and the ability to learn new exercises by the user.

The increasing public interest in fitness trackers as consumer products, in addition to their potential uses in health care necessitate the presence of proper evaluations for all the aspects of these devices. The huge number of devices available on market and the speed in which new devices with new features are released, makes it very difficult to come up with a proper evaluation for market products that won't quickly turn obsolete especially when taking longer device usage-time as a factor in evaluation. In (Shafaee et al., 2014), we introduced a new framework for evaluating fitness trackers based on the public opinion expressed through Amazon.com product reviews. This framework follows an aspect-based sentiment analysis approach that dynamically identifies aspects of every fitness tracker and evaluates all sentiment-bearing mentions of an aspect in all Amazon.com reviews of the tracker.

In this paper, we present an extensive evaluation of the top three market devices from manufacturers that together dominated over 97% of the fitness trackers market in 2013 according to NPD Group[4], namely;

---

[1] http://www.samsung.com/us/globalinnovation/

[2] Electrocardiogram; spelled with a 'K' because in German it is spelled *Elektrokardiogramm*

[3] https://amiigo.com/

[4] https://www.npd.com/latest-reports/consumer-technology-reports/

Fitbit One[5], Jawbone Up[6] and Nike+ Fuelband[7]. It is worth mentioning that the choice of these three products is strictly for their leading position in the consumer market and that there are no limitations by our framework to include any other fitness/health tracker available for sale on Amazon.com. Not only an answer to the general question *"which is the best fitness tracker?"* is given in this work, but also to the more personal question *"which fitness trackers suits me best?"* is given in Section 5.4. A personalized recommender system for suggesting the best fitness tracker for each individual based on her needs and preferences is thus presented.

The rest of this paper is organized as follows. Section 2 discusses the related work for this paper. Section 3 introduces the devices and the dataset used for the evaluation. An overview of the back-end sentiment analysis framework used as basis for the presented sentiment summarizer is given in Section 4. Results and analysis of the Evaluation are presented in Section 5, and finally the conclusions and future work are discussed in Section 6

2 RELATED WORK

The sentiment analysis framework introduced in Section 4 is further detailed and evaluated in (Shafaee et al., 2014). The system components are proven to top state-of-the art alternatives. Studies that address the accuracy of fitness trackers are relatively few. A study (Dannecker KL, 2014) aimed at evaluating the accuracy of energy expenditure estimation of consumer physical activity monitors found out that Fitbit sensors tend to significantly underestimate energy expenditure. In (Guo et al., 2013), an extensive physical testing for several fitness trackers and pedometers including Nike+ Fuelband and Fitbit One is presented. The study shows that Fitbit One is the most accurate in counting steps, which verifies the results presented in Section 5.2.1.

3 DEVICES AND DATASET

3.1 Devices

In this section a brief introduction to the three devices used in evaluation is given. The devices considered are shown in Figure 1.

---

[5]http://www.fitbit.com/one

[6]https://jawbone.com/up

[7]http://www.nike.com/us/en_us/c/nikeplus-fuelband

Figure 1: From left to right: Fitbit One, Jawbone Up, Nike+ Fuelband.

3.1.1 Fitbit One

Fitbit one is a clip-on tracker that was announced on September 17, 2012. It features a digital display and it is the first tracker to sync through Bluetooth 4.0. Fitbit One supports Android, iOS and Windows Phone smartphone operating systems. It can record several daily activities including number of steps taken, distance travelled on foot, number of floors climbed, calories burned, vigorously active minutes, as well as quality of sleep. It also incorporates a "silent alarm" that wakes the user up through gentle vibrations. The launch price for Fitbit One was 99.99 U.S. dollar.

3.1.2 Jawbone Up

Jawbone up was initially announced in November 2011. Due to manufacturing problems, the product was relaunched in November 2012. The Up comes in the form of a waterproof wristband. It can track number of steps taken, distance travelled on foot, calories burned as well as quality of sleep. It incorporates a vibration alarm feature and communicates to many 3rd party lifestyle and fitness apps and services. Jawbone Up supports Android and iOS through a native smartphone app. The launch price for Jawbone Up was 129.99 U.S. dollar.

3.1.3 Nike+ Fuelband

Nike+ Fuelband is a wristband tracker that was announced on January 19, 2012. Fuelband uses a set of LEDs as a screen. It tracks the number of steps taken and calories burned, it displays time and computes a proprietary measure of fitness activity, NikeFuel. Users set their daily goal in terms of NikeFuel, and the band displays the progress of the user in achieving their goal through an array of colored LEDs. The Fuelband syncs wirelessly to smartphones and both Android and iOS are supported through a native app. The launch price for Nike+ Fuelband was 149.99 U.S. dollar.

3.2 Dataset

For the evaluation, a total of 3,241 Amazon.com reviews of the three fitness trackers Fitbit One (FO), Jawbone Up (JU) and Nike+ Fuelband (NF) are used. The distribution of the reviews (REV), sentences (SEN), aspect mentions (MEN) and sentiment-bearing aspect mentions (SBM) among the three devices is shown in Table 1. The significantly lower number of reviews available for Nike+ Fuelband does not necessarily reflect its market share with respect to the two other trackers. It is probably due to sales through Nike's channels rather than Amazon.com.

Table 1: Details about the dataset used for evaluating the devices.

Device	REV	SEN	MEN	SBM
FO	2,504	15,154	7,682	6,372
JU	696	4,397	2,138	1,645
NF	41	246	818	612
Total	3,241	19,797	10,638	8,629

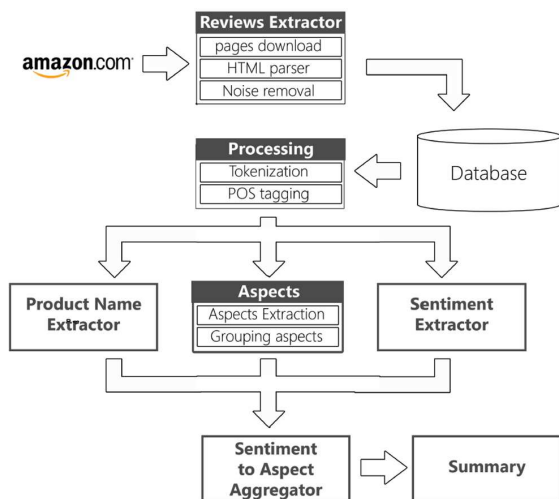## 4 OVERVIEW OF THE SENTIMENT ANALYSIS SYSTEM USED

Figure 2: System Overview.

A general overview listing the main components of the sentiment analysis system used for the evaluations is shown in Figure 2. For each product, its name or Amazon ID(s) are fed into the system. Since Amazon no longer offers reviews through its API, a script is utilized to extract reviews using regular expressions. Any noise caused by mismatched HTML tags is removed and the reviews are then inserted into a database. The Stanford CoreNLP (Manning et al., 2014) is then used to extract sentences in each review, tokenize them and get the part-of-speech and base form for each token. The output of the previous step is then fed into three other components, namely, product name extractor, aspect extractor and sentiment extractor which are defined below.

- **Product-Name Extractor:** The product name extractor has two main aims. First, it automatically extracts the names of competing products. Second, it identifies mentions of the given product or competing products in text. For example, if an author says: "I love my Fitbit One. I got it after I broke my Ultra 2 months ago.", the system identifies that the first sentence discusses "Fitbit One", whereas the latter discusses "Fitbit Ultra". This component is essential to assign the sentiment to the proper product and is useful in the other components.

- **Aspect Extractor:** This component extracts the aspects of the product that reviewers commented on. For fitness trackers, the aspects can be price, customer service, steps counter, mobile apps, etc. Aspects are classified into static aspects and dynamic aspects. Static Aspects are features that are common to all products such as price and customer service while dynamic aspects are features that may differ from a product to another. For fitness trackers, heart rate monitoring is an example of a dynamic aspect since not all fitness trackers monitor heart rate.

- **Grouping Aspects into Categories:** After identifying the aspects of the product, sentences discussing each aspect are extracted. Given that an aspect may be discussed by different people using different terms, all the different mentions of the same aspect are grouped. To illustrate, consider the direct and indirect mention of the "price" aspect in these two sentences: "Nike+ Fuelband is quite *expensive*." and "The Fuelband has a *high price tag*!"

- **Sentiment Extractor:** The sentiment extractor component aims at detecting the opinion of a whole sentence or part of it. That is, given a phrase, it returns whether it is positive, negative or neutral. For instance, "I love this device" is positive, "This device is really useless" is negative and "I bought this device yesterday" is neutral.

The next step is to assign the sentiment to the proper aspect. Finally, the data collected from the

Table 2: Fitbit One Results.

Aspect	Positive	Negative	Neutral	Total
Steps Counter	605	183	492	1280
Sleep Assessment	570	199	373	1142
Syncing	408	146	263	817
Calories Burned	315	113	245	673
Floors Climbed	276	99	267	642
App Functionality	332	114	182	628
Website	332	85	176	593
Battery	206	100	201	507
Price	186	102	124	412
Pocket/Clipping	109	67	112	288
Customer Service	148	50	60	258
Distance Traveled	90	23	88	201
Badges	71	12	32	115
MyFitnessPal	38	9	21	68
Activity Indicator (Flower)	36	8	14	58
Total	3722	1310	2650	7682

sentiment analysis of the reviews for each product is presented through an interactive web-based summarizer that provides the user with statistics and insights about the products. Further details on the components presented are discussed in (Shafaee et al., 2014).

5 EVALUATION RESULTS AND ANALYSIS

In this section, the results of the sentiment analysis process introduced in Section 4 applied to the three devices chosen for this study are presented. Table 2, Table 3 and Table 4 show the aspects detected by the system and the distribution of the sentiments among these aspects for Fitbit One, Jawbone Up and Nike+ Fuelband respectively. Before discussing the output of the evaluation, a precision/recall analysis of the system is first presented to estimate the credibility of the evaluations following in this section. For this sake, human annotators were asked to annotate all sentiments in 106 Amazon reviews distributed almost equally among the three products. The results for our system were compared to two state-of the-art supervised methods, namely Support Vector Machine (SVM) trained on unigrams and Naive Bayes classifiers trained on unigrams(Pang et al., 2002). Figure 3 shows that our system has the highest recall with very high precision for the positive sentiment class. Similarly, Figure 4 shows our system has significantly higher precision and a fairly good recall when compared to the other approaches for the negative sentiment class. A big advantage of the proposed system is that it is lexicon-based which makes it possible to

generalize to different domains. This is not the case with the supervised approaches which require a large set of annotated data and can therefore poorly generalize.
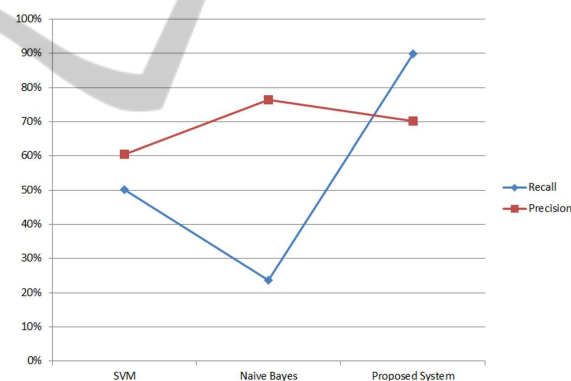

Figure 3: Recall and precision of sentiment detection for the positive sentiment class.

5.1 Controversial Aspects

Regardless of the sentiment, the frequency of aspect mentions in reviews is very significant, especially for evaluating distinguished aspects that set a product apart from its market competitors. For example, the NikeFuel activity measure introduced in the Nike+ Fuelband is the most discussed aspect for this device with 22.7% of the Fuelband's aspect mentions discussing this unique aspect. In contrast, the flower activity indicator in the Fitbit One, which resembles NikeFuel for the Fuelband, takes a share of 0.7% only from the Fitbit's aspect mentions and it is
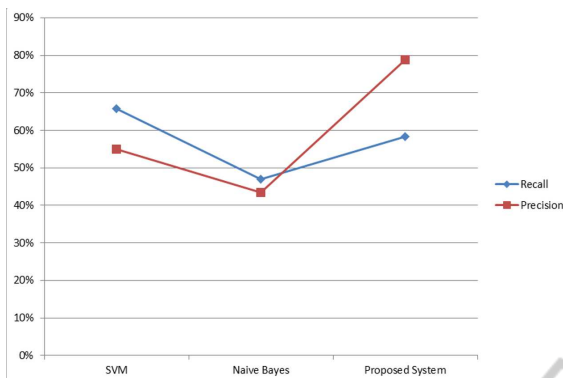
Figure 4: Recall and precision of sentiment detection for the negative sentiment class.

the least discussed aspect for the device. This indicates that Nike's distinguished activity measure has successfully captured the interest of consumers where as Fitbit falls short in this regard. The Power Nap aspect of Jawbone Up, which is a feature that calculates the optimal nap duration for the user and wakes them up through gentle vibrations, is barely discussed with 0.7% of aspect mentions despite being an innovative and positively received feature. This signifies that the usability of this feature can be questioned. However, Sleep assessment, in general, has proven to be an important feature to users grabbing over 20% of Jawbone's aspect mentions and about 15% of Fitbit's aspect mentions. This indicates that Nike is missing a highly demanded feature by not including it in their device. Also, Fitbit's ability to compute floors climbed holds over 8% of its aspect mentions almost equal to the share of "Calories Burned" feature and thus it also represents a feature demanded by consumers.

5.2 Static Aspects Evaluation

Static aspects describe essentially the basic features and capabilities of a fitness tracker. In this section, we assess the public satisfaction of the performance-related aspects in all three devices. Namely: *Steps Counter, Distance Traveled and Calories Burned*. We point here, that similar insight and figures can be exported by the sentiment summarizer for all the other aspects. However, due to space restrictions and to avoid repetitions, detailed comparison for the other aspects are omitted. Still, useful insights for each device are given in Section 5.3.

5.2.1 Steps Counter Evaluation

All three devices seem to be satisfactory to consumers with respect to steps counting as shown in Figure 5

despite a significant advantage for the Fitbit One. This actually supports the physical testing of accuracy of several fitness trackers including Fitbit One and Nike+ Fuelband by (Guo et al., 2013) in which the One outperformed all other devices.
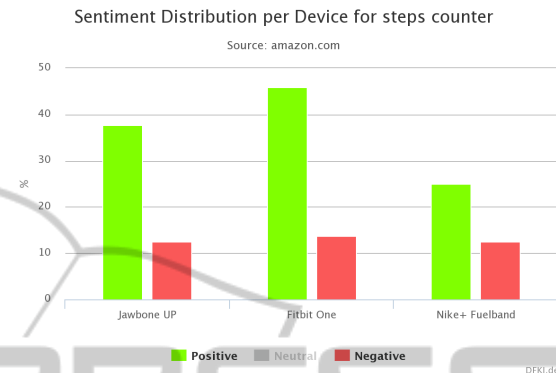

Figure 5: Results of the public sentiment evaluation for aspect "Steps Counter" as displayed by the sentiment summarizer.

5.2.2 Distance Travelled Evaluation

The good performance in counting the steps by the devices leads to a good estimation of the distance travelled by users as shown in Figure 6. Since all three devices do not include location sensors, the distance value is probably calculated through estimations based on the individual's personal data provided to the device during the set up process.
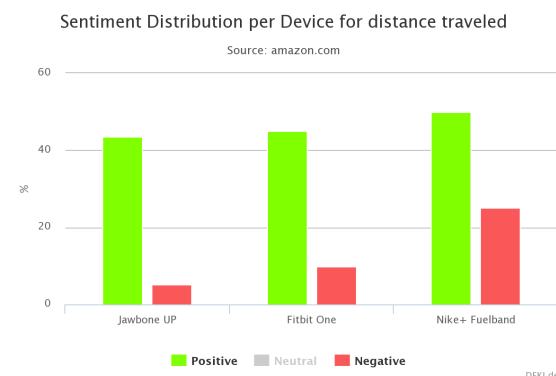

Figure 6: Results of the public sentiment evaluation for aspect "Distance Travelled" as displayed by the sentiment summarizer.

5.2.3 Calories Burned Evaluation

In a similar manner, the devices estimate calories burned based on the performance of the user and his/her entered physical data (e.g. height, age, weight,

Table 3: Jawbone Up Results.

Aspect	Positive	Negative	Neutral	Total
Sleep Assessment	225	82	128	435
Battery	113	113	123	349
App Functionality	159	57	95	311
Steps Counter	90	30	97	217
Syncing	78	65	66	209
Price	63	55	30	148
Customer Service	51	31	45	127
Alarm	51	15	24	90
Calories Burned	38	13	27	78
Food Logging	28	18	15	61
Website	19	5	16	40
Distance Traveled	17	2	16	35
Powernap	6	2	8	16
Pocket	7	4	3	14
Activity Logging	5	1	2	8
Total	950	493	695	2138

Table 4: Nike+ Fuelband Results.

Aspect	Positive	Negative	Neutral	Total
NikeFuel	70	49	67	186
Price	41	29	38	108
Battery	35	21	41	97
Calories Burned	33	21	38	92
Steps Counter	29	22	31	82
App Functionality	38	15	22	75
Syncing	29	20	18	67
Website	19	10	23	52
Customer Service	10	13	9	32
Distance Traveled	8	6	13	27
Total	312	206	300	818

gender, etc.). Both Fitbit One and Jawbone Up seem to be equally satisfactory for users estimating the calories burned where as the Fuelband has a significantly less positive-to-negative mention ratio compared to the two other devices as shown in Figure 7.

5.3 Further Discussion and Insights

In addition to the statistics provided, the sentiment summarizer allows for easy tracking of positive, negative and neutral mentions of each aspect of every device through out the reviews. For example (Figure 8), a user can request all the negative mentions of 'Customer Service' for Jawbone Up and the summarizer will return a list of all sentences with these negative mentions sorted by the degree of negativity estimated by the system. This gives a new dimension in the evaluation by highlighting pros and cons based on the experience of customers. Many useful conclusions can
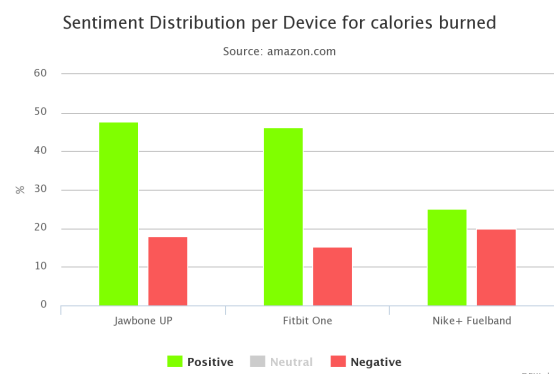

Figure 7: Results of the public sentiment evaluation for aspect "Calories Burned" as displayed by the sentiment summarizer.

be drawn from people's opinion. Some of them are given below.

I am glad I got to try this product for free through my job rather than wasting over a hundred dollars on an item with poor build and poor customer service.

Terrible customer service.

I ran through all the steps I'd learned from customer support earlier, but the problem continued until the bracelet actually crashed my iPhone.

Very disappointed with the Jawbone customer service!

Customer service doesn't seem very responsive!

However, there is no manual and customer support is unresponsive.

I have never, not once, been told by any customer support agent ANYWHERE that they couldn't exchange an item for me or help me with a problem just because I didn't purchase it directly from the company.

Figure 8: Snapshot of a part of the list of negative *Customer Service* aspect mentions for Jawbone Up as shown in the sentiment summarizer.

5.3.1 Fitbit One

Many proponents of Fitbit commented on how it helped them greatly pay attention to their sleep quality and how it encouraged better sleeping habits. The ratio between the number of users in favor of the sleep assessment made by the tracker to those against it is almost 3:1. Some customers however, found the sleeping band to be uncomfortable or the sleeping data to be inaccurate. One of them said: *"Tossing in the middle of the night or pushing covers off your body will register as waking up, and watching an hour or two of television without moving your hand before falling asleep will count as sleep time. Unfortunately, this makes the sleep tracker a good idea that is not very accurate."*

Some reviewers commented that it holds a good charge (e.g., charging it only on weekends) and some were pleased that it emails them a reminder when the charge is low. Quoting a user *"The charging cord and Bluetooth dongle are better executed than the stand used in the previous version, and it's MUCH easier to travel with."*. However, some said that their Fitbits did not hold a charge for more than 8-12 hours and others experienced the total death of the device.

Many users find setting up the Fitbit One and syncing the data very easy and smooth. They also love syncing the Fitbit with many third-party apps especially to MyFitnessPal app. In fact, the system identified MyFitnessPal as one of the aspects of Fitbit because people comment a lot on syncing their Fitbits to it. For the Fitbit One app itself, some people claim that it is buggy for most Androids. Others complained that it requires a dongle to sync to PC or MAC saying that the dongle is very small and can be easily lost. Some could not pair it to Android. A frequent traveler said that the time shown after syncing is not accurate.

Although people are pleased that it is small and light, many people lost it because, according to them, it can easily fall out of the holder and the holder does not attach firmly to clothes. Therefore, they highly recommend having it inside the pocket rather than clipped onto the pocket or belt. Despite some who are motivated by the growth of the "activity-progress flower", many others believe that Fitbit should change this visualization.

5.3.2 Jawbone Up

The most popular feature of Jawbone Up is sleep assessment. Most of the customers who reviewed this device expressed how they like its accompanying features, namely the silent vibrating alarm, Power Nap which wakes a user up at the optimal time, and the Idle Alert which reminds users to move when they have been inactive for too long. Nonetheless, many people complained of having to change it manually to sleep mode before sleeping and changing it back to awake mode later, which one can forget. Some customers also questioned the accuracy of the sleep information the band provides.

A major problem many users faced is the complete death of the band after a period of use. Nonetheless, many users said that the customer service replaced their bands. On the other hand, some described the customer service as poor and one of the complaints said that they could not exchange the item or help with its problems because it was not bought directly from the Jawbone company.

Many people recommend calibrating the band to record the distance accurately. Some sentences analysed by the summarizer stated that wearing the Jawbone on ankle is said to increase the accuracy of the band, yet, some commented that it is uncomfortable.

Some customers said that they did not experience any problems with Syncing. However, the Jawbone UP does not sync wirelessly which made some complain about having to take the band cap off and inserting the band in the headphone's jack of the phone every time. Last but not least, the Jawbone Up app was generally positively rated by most consumers.

5.3.3 Nike+ Fuelband

NikeFuel score is perhaps the most remarkable feature of the Fuelband being, by far, the most reviewed aspect of the band as shown in Table 4. By reading some of the sentences discussing this aspect, the Fuel score seems very motivating for many. However, some are skeptical of how it is calculated. A user said *"I've noticed that the Fuel really adds up when running and walking but not as much with other workouts that are way more intense than a jog on the treadmill.".* Another said *"Of course, if you shake your wrist while on the elliptical or jogging, you can increase your points by 50%."*. Some people said that it is very motivational, however, they did not recommend it for people who want to accurately track their steps. With respect to syncing, some people found syncing over the phone or USB to be smooth and easy. On the other hand, it failed to sync with some and was described as slow. Some users complained that the screws in the band rust. One suggested using a fingernail polish to stop the rust. Many users liked the accompanying app which was described as motivational, easy-to-use and well-designed. In fact, the number of positive reviews on the Nike+ Fuelband app is more than twice the negative reviews on it. Yet, some were disappointed that it did not come with an Android app.

5.4 Personalized Fitness Tracker Recommender

Given the vast knowledge the presented sentiment summarizer acquired by analysing the sentiments in Amazon.com reviews of fitness trackers, the second step is to utilize this knowledge to help each potential customer choose the device that suits her needs best based on her own preferences. Figure 9 shows the user interface for the device recommender in the sentiment summarizer. A user can select using 5-value sliders, how important an aspect is to her, and the recommender will dynamically upgrade score gauges for each device thereby helping each customer making the right purchase decision based on all the acquired knowledge of owners of these devices.

The scores are computed using Formula 1; where d denotes a device, k denotes the total number of aspects for which the user have assigned a non-zero weight using the slider, a_i denotes a single aspect ($i \in [1, k]$), $q_{a_i} \in [0, 4]$ denotes the weight assigned by the user to aspect a_i, $pos(a_i)$ and $neg(a_i)$ denote the number of positive (res. negative) mentions of the aspect a_i for device d, and A_d denotes the set of all aspects for device d. Intuitively, when the user gives a

non-zero weight for a certain aspect using the slider, this aspect is added to the set of user-requested aspects. For each aspect a_i of these k aspects, the system checks if it is included in the aspects of the considered device. If yes, a value $\frac{pos(a_i)-neg(a_i)}{pos(a_i)+neg(a_i)} \in (-1, 1)$ is returned indicating how recommended (res. unrecommended) the device is when considering this aspect alone. Otherwise, if the aspect requested by user is not available for the device, -1 is returned indicating that the device is not recommended at all when considering this aspect. The result is adjusted for each aspect based on a user-selected weight q_{a_i}. The final score $Score_d$ is then computed as the normalized mean for the returned weighted values for each aspect.

Figure 9: Snapshot of the personalized device recommender functionality in the sentiment summarizer.

$$Score_d = \frac{100}{k} \sum_{a_i=1}^{k} \frac{q_{a_i}}{4} \times \begin{cases} \frac{pos(a_i)-neg(a_i)}{pos(a_i)+neg(a_i)} & if \quad a_i \in A_d \\ -1 & otherwise \end{cases}$$

(1)

6 CONCLUSIONS AND FUTURE WORK

In this paper, the novel fitness-trackers' evaluation approach presented in (Shafaee et al., 2014) is demonstrated to evaluate three market leading fitness trackers. The approach relies on aspect-based sentiment analysis of Amazon.com reviews of these products. The results of the evaluation are presented to the user in an easy-to-use web-based summarizer. Statistics and insights resulting from the evaluation are presented in Section 5 where also a personalized fitness tracker recommender system is presented. In addition to the ongoing research on enhancing the underlying algorithms used in the back-end sentiment analysis system introduced in Section 4, our efforts are also oriented on automating the whole process starting from fetching the Amazon.com reviews until reaching an always up-to-date summarizer that adds

even new measures to its insights, including sentiment variation over time, and per different countries.

ACKNOWLEDGEMENTS

REFERENCES

BCCResearch (2014). Wearable computing: Technologies, applications and global markets. Technical report, BCC Research.

Dannecker KL, Sazonova NA, M. E. S. E. B. R. (2014). A comparison of energy expenditure estimation of several physical activity monitors. *Medicine and science in sports and exercise*.

Guo, F., Li, Y., Kankanhalli, M. S., and Brown, M. S. (2013). An evaluation of wearable activity monitoring devices. In *Proceedings of the 1st ACM International Workshop on Personal Data Meets Distributed Multimedia*, PDM '13, pages 31–34, New York, NY, USA. ACM.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 55–60.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

Shafaee, A., Issa, H., Agne, S., Baumann, S., and Dengel, A. ("2014"). Aspect-based sentiment analysis of amazon reviews for fitness tracking devices. In *PAKDD Workshops*, pages "50–61".