Sensing Emotions in Text Messages:
An Application and Deployment Study of EmotionPush

Shih-Ming Wang[1]   **Chun-Hui Li**[1]   **Yu-Chun Lo**[1]
Ting-Hao (Kenneth) Huang[2]   **Lun-Wei Ku**[1]
[1] Academia Sinica, Taipei, Taiwan.
{ipod825, iamscli.tw}@gmail.com, howard.lo@nlplab.cc, lwku@iis.sinica.edu.tw
[2] Carnegie Mellon University, Pittsburgh, PA, USA. tinghaoh@cs.cmu.edu

Abstract

Instant messaging and push notifications play important roles in modern digital life. To enable robust sense-making and rich context awareness in computer mediated communications, we introduce *EmotionPush*, a system that automatically conveys the emotion of received text with a colored push notification on mobile devices. EmotionPush is powered by state-of-the-art emotion classifiers and is deployed for Facebook Messenger clients on Android. The study showed that the system is able to help users prioritize interactions.
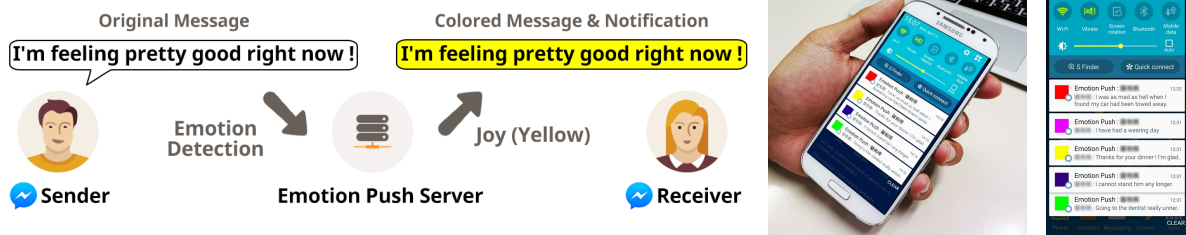
1   Introduction

Figure 1: System overview and user scenario of EmotionPush. The server detects the emotion of the message and sends a colored notification to indicate the sender's emotion on the receiver's device.

Text-based communication plays a big part in computer mediated communications. With the advent of mobile devices, instant messaging and push notifications have become integral to modern digital life. However, text-based chatting is limited in both expressing emotions and building trust amongst users. Text-only chatting has been shown to result in worse communications and trust levels than face-to-face, video, and audio virtual communications (Bos et al., 2002). One study also demonstrated that knowing other people's emotions is crucial for collaboration, yet is surprisingly challenging for users via computer mediated communications (Eligio et al., 2012).

In response, previous works have attempted to automatically visualize emotions on text-based interfaces by visualizing the emotion dynamics in a document (Liu et al., 2003), providing haptic feedback via wearable equipment (Tsetserukou et al., 2009), or changing font sizes according to recognized emotions (Yeo, 2008). However, these explorations were primarily developed based on rule-based emotion detectors, which were shown to perform significantly worse than machine-learning algorithms (Wu et al., 2006). On the other hand, some researchers proposed to add new features, such as kinetic typography (Bodine and Pignol, 2003; Forlizzi et al., 2003; Lee et al., 2006; Lee et al., 2002), affective buttons (Broekens and Brinkman, 2009), and two-dimensional representations (Sánchez et al., 2005; Sánchez et al., 2006), on top of traditional chatting interfaces to allow users express emotions. Others studies have attempted to incorporate the user's body signals, such as fluctuating skin conductivity levels (DiMicco et al., 2002), thermal feedback (Wilson et al., 2016), or facial expression (El Kaliouby and
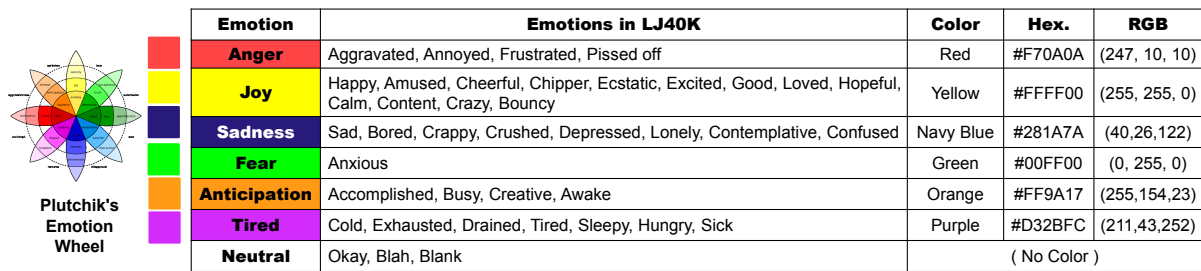
Emotion	Emotions in LJ40K	Color	Hex.	RGB
Anger	Aggravated, Annoyed, Frustrated, Pissed off	Red	#F70A0A	(247, 10, 10)
Joy	Happy, Amused, Cheerful, Chipper, Ecstatic, Excited, Good, Loved, Hopeful, Calm, Content, Crazy, Bouncy	Yellow	#FFFF00	(255, 255, 0)
Sadness	Sad, Bored, Crappy, Crushed, Depressed, Lonely, Contemplative, Confused	Navy Blue	#281A7A	(40,26,122)
Fear	Anxious	Green	#00FF00	(0, 255, 0)
Anticipation	Accomplished, Busy, Creative, Awake	Orange	#FF9A17	(255,154,23)
Tired	Cold, Exhausted, Drained, Tired, Sleepy, Hungry, Sick	Purple	#D32BFC	(211,43,252)
Neutral	Okay, Blah, Blank	(No Color)		

Plutchik's Emotion Wheel

Figure 2: Visualizing emotion colors with Plutchik's Emotion Wheel. The 40 emotion categories of LJ40K are compacted into 7 main categories, each has a corresponding color on the emotion wheel.

Robinson, 2004), in instant messaging applications. However, after decades of research, these features are largely absent in modern instant messaging clients.

In this paper, we introduce **_EmotionPush_**, a system that displays colored icons on push notifications (as shown in Figure 1) to signal emotions conveyed in received messages. EmotionPush is powered by machine learning technologies with state-of-the-art performances. Built on top of the long-lasting development of emotion detection, we applied the techniques to a real-world chatting environment to determine how well the system works for individual users. Our contribution is two-fold: 1) We created EmotionPush, the first system powered by modern machine-learning emotion classification technology to convey emotions for instant messages, and 2) we deployed the system on a widely-used instant messaging client on mobile devices, Facebook's Messenger, to examine the feasibility of the emotion feedback.

2 EmotionPush System

Similar to most mobile apps, EmotionPush adopted a client-server architecture (as shown in Figure 1.) When the user (receiver) receives a message via the instant messaging client, our system uses the text of the message to recognize its corresponding emotion, and then notifies the user (receiver) via push notification with a colored icon on his/her mobile device. We developed the EmotionPush _client_ as an Android application[1], specifically for Facebook's Messenger (https://www.messenger.com/). The screen shot and user scenario of the app are shown in Figure 1. The EmotionPush _server_ was implemented as a stand-alone web server powered by pre-trained emotion classification models.

Visualizing Emotions   EmotionPush uses 7 colors to represent 7 emotions, as shown in Figure 2. This schema was designed as follows: First, we focused on emotions commonly connected with life events, unlike benchmarks such as (Nakov et al., 2016) which typically focus on general social media data. To simplify the mapping between emotions and text, we also decided to apply a _categorical representation_ (e.g. _Anger_, _Joy_, etc.) (Klein et al., 2002) of emotions instead of a dimensional representation (valence, arousal) (Sánchez et al., 2006). Second, we utilized the emotion categories and data provided in _Live-Journal_ (http://www.livejournal.com/). LiveJournal is a website where users post what they feel and tag each post with a corresponding emotion. The _LJ40k_ corpus (Leshed and Kaye, 2006), a dataset that contains 1,000 blog posts for each of the 40 most common emotions on LiveJournal, was adopted to learn which emotions we should watch for in EmotionPush and to train the emotion classifiers. Finally, to reduce users' cognitive load, the original 40 emotions were compacted into 7 main emotions according to Plutchik's Emotion Wheel color theme (Plutchik, 1980), as shown in Figure 2.

Emotion Classification & Evaluation   EmotionPush's 7 classifiers were trained on the LJ40k dataset with 7 compacted emotion labels. Each classifier is a binary classifier that indicates if the current message belongs to one of the 7 compacted emotions. The message was inputted into each classifier to obtain the probability of each compacted emotion label. Then the label of the highest probability was selected as the predicted emotion label of the current message.

To compare the performance of our approach with that of previous works on the LJ40k dataset (Yang and Liu, 2013), we replicated our classification method and features to predict the original _40_ emotions

---

[1]EmotionPush is available at Google Play: https://play.google.com/store/apps/details?id=tw.edu.sinica.iis.emotionpush
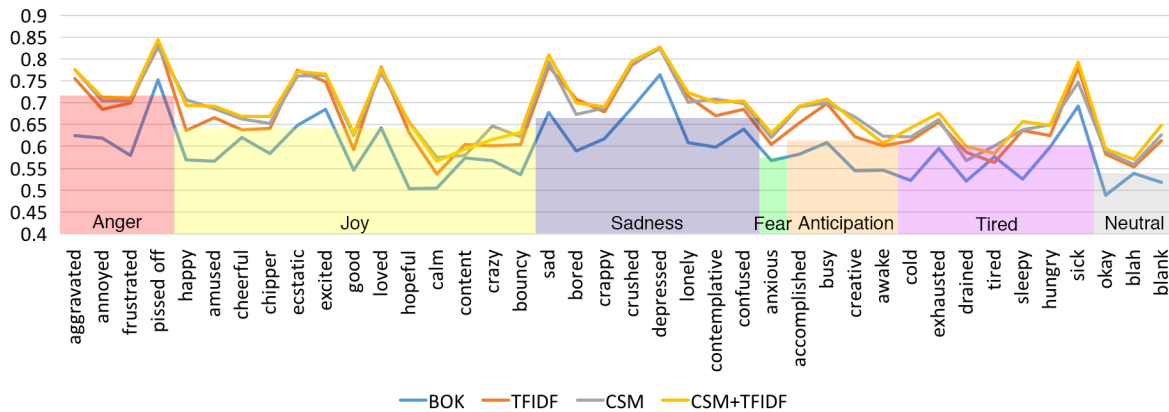
Figure 3: AUC scores of the binary classifiers corresponding to the 40 LiveJournal emotions and the 7 major emotions in EmotionPush.

in LJ40k. Forty binary-class classifiers (one emotion each) were developed by using LibSVM (Fan et al., 2008) with a radial basis function kernel. We chose to develop 40 classifiers instead of one 40-class classifier to a) better compare our results to Yang and Liu (2013), b) achieve better performance, and c) extend EmotionPush to a multi-labeling system in the future. To form a balanced training set for each emotion, we randomly selected 800 posts from LJ40K as positive examples and 800 posts of the other 39 emotions as negative examples. Aware of various features proposed for affect (Ku et al., 2009; Ku et al., 2011; Balahur et al., 2014; Poria et al., 2014; Tang et al., 2014; Xu et al., 2015), we used a 300-dimension word vectors trained on Google News (Mikolov et al., 2013) (https://code.google.com/archive/p/word2vec/) to represent each post. The model's parameters were tuned via a 10-fold cross-validation process.

The evaluations were performed on the held-out testing set that contains 8,000 posts (200 posts for each emotion). The AUC, the area under the receiver operating characteristic curve, was calculated. For the 40 emotions, our classifiers achieved an average AUC of 0.6788, which was comparable to the state-of-the-art performance, 0.6851, reported by (Yang and Liu, 2013). Figure 3 shows the performance of each emotion using different features. We observed that classifiers performed worse on *Blank*, *Okay*, *Drained*, and *Sleepy*. These low-performance emotions can be roughly classified into two categories: 1) *vague emotions*, such as *Blank* and *Okay*, are difficult to model, as people tend to interpret them differently, and 2) *similar emotions*, such as *Drained* and *Sleepy*, tend to overlap significantly, and thus hinder us from distinguishing them.

Furthermore, the colored bars in Figure 3 show the classification performances for the 7 compacted major emotions for EmotionPush. The classifiers of compacted emotions *Joy*, *Sadness* and *Anger* performed best among all 7 emotions, while *Neutral* and *Fear* performed worse, which might be because these two compacted emotions were made up of fewer LiveJournal emotions than others, as shown in Figure 2. This not only resulted in the lack of training data, but also brought in errors as these few emotions (including *Anxious*, *Okay*, *Blah*, and *Blank*) performed comparably less satisfactory (see Figure 3).

3 Deployment Study

In this study, we aimed to test whether EmotionPush can change the priority of interactions in instant messages on mobile devices. Therefore, we deployed EmotionPush to Google Play, and recruited 8 native English speakers who frequently used Facebook's Messenger app. The experiment lasted 12 days. We investigated the effect of EmotionPush by turning the color feedback off (for the first 5 days, noted as the *first week*) and on (for the latter 7 days, *second week*). We then analyzed changes in participants' priorities of reading and responding to messages.

During the entire study, we collected 6,288 messages in total, 3,844 read counts (first read of a message sequence) and 3,769 response counts (first response). The overall average score obtained from participants was above the average (2.375 over 4) for the question "EmotionPush can predict emotion colors correctly." Moreover, participants did not think wrongly predicted emotions would harm their chatting
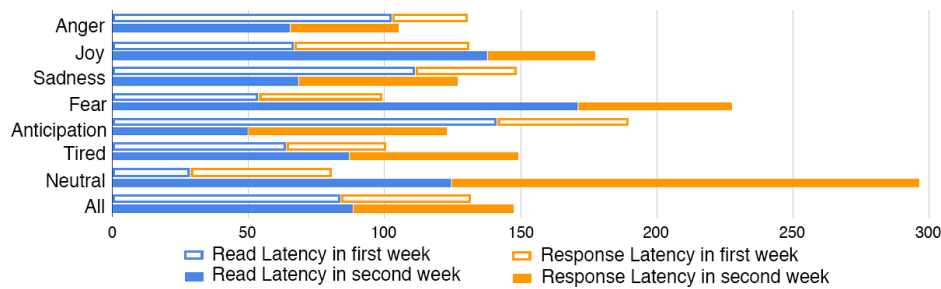
Figure 4: The results of two-week deployment study of EmotionPush. This chart shows the users' read and response latencies when color feedback is on (first week) and off (second week), respectively.

experience (average score 1.375 over 4).

From this study, we found that EmotionPush helps **prioritize interactions**. The average read latencies and response latencies for the first week and the second week are shown in Figure 4. We can observe that after the emotion colors were pushed, user's behavior changed accordingly. Especially for *Joy*, *Anger*, and *Sadness*, the priority of the instant message interactions changed in two interesting directions: 1) *Joy* (blue bar: with an increasing read latency from first week to second week) was read more slowly while *Sadness* and *Anger* (blue bar: with a decreasing read latency) was read more quickly, and 2) *Joy* was responded to more quickly (orange bar: with a decreasing response latency) while *Sadness* and *Anger* were responded to more slowly (orange bar: with an increasing response latency). The decreasing read latency of *Anger* ($p = 0.0019$) and the increasing response latency of *Sadness* ($p = 0.0486$) are significant. This might reveal that messages that are not urgent (*Joy*) can be put aside and read later, while participants are willing to read urgent messages (*Sadness* and *Anger*) earlier. On the other hand, users could casually respond to less urgent messages (i.e., with little thought), whereas urgent messages require more thought.

An interesting thing to mention is the changes of the response latency. Logically speaking, as participants read messages, they should realize the messages' emotions so that their response latency would not change after pushing emotion colors. However, we observed changes in Figure 4, and there is even a significant difference of the response latencies for *Sadness*. This suggests that the color feedback is not only notifying users, but also influencing their process of composing a response.

Overall, the feedback for using EmotionPush was positive. 78% of participants thought it is a good idea to add the feature of EmotionPush to Facebook's Messenger, and the other 22% wanted to add this feature eventually but just not immediately to wait for a better user interface and a better prediction performance for some emotion categories.

4   Conclusion and Future Work

We introduced EmotionPush, a system that automatically recognizes and pushes emotions of instant messages and visualizes them to the end-user, which enables the emotion sensing ability on messages and enriches the information in the communications. We believe this research can help us gain more insight into the effect of reinforcing the emotion sensing of robots. In the future, we plan to add this emotion sensing function in the message composing process and discuss the quality of conversations.

References

A. Balahur, R. Mihalcea, and A. Montoyo. 2014. Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications. *Computer Speech & Language*, 28(1):1–6.

K. Bodine and M. Pignol. 2003. Kinetic typography-based instant messaging. In *CHI'03 EA*. ACM.

N. Bos, J. Olson, D. Gergle, G. Olson, and Z. Wright. 2002. Effects of four computer-mediated communications channels on trust development. In *Proc. CHI'02*. ACM.

J. Broekens and W. Brinkman. 2009. Affectbutton: Towards a standard for dynamic affective user feedback. In *Proc. ACII'09*. IEEE, September.

J. M. DiMicco, V. Lakshmipathy, and A. T. Fiore. 2002. Conductive chat: Instant messaging with a skin conductivity channel. In *Proceedings of Conference on Computer Supported Cooperative Work*.

R. El Kaliouby and P. Robinson. 2004. Faim: integrating automated facial affect analysis in instant messaging. In *Proceedings of the 9th international conference on Intelligent user interfaces*, pages 244–246. ACM.

U. X. Eligio, S. E. Ainsworth, and C. K. Crook. 2012. Emotion understanding and performance during computer-supported collaboration. *Computers in Human Behavior*, 28(6).

R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin. 2008. Liblinear: A library for large linear classification. *The Journal of Machine Learning Research*, 9.

J. Forlizzi, J. Lee, and S. Hudson. 2003. The kinedit system: affective messages using dynamic texts. In *Proc. CHI'03*, pages 377–384. ACM.

J. Klein, Y. Moon, and R. W. Picard. 2002. This computer responds to user frustration:: Theory, design, and results. *Interacting with computers*, 14(2):119–140.

L.-W. Ku, T.-H. Huang, and H.-H. Chen. 2009. Using morphological and syntactic structures for chinese opinion analysis. In *Proc. EMNLP '09*, EMNLP '09, pages 1260–1269, Stroudsburg, PA, USA. ACL.

L.-W. Ku, T.-H. K. Huang, and H.-H. Chen. 2011. Predicting opinion dependency relations for opinion analysis. In *IJCNLP*, pages 345–353.

J. C. Lee, J. Forlizzi, and S. E. Hudson. 2002. The kinetic typography engine: an extensible system for animating expressive text. In *Proc. UIST'02*, pages 81–90. ACM.

J. Lee, S. Jun, J. Forlizzi, and S. E. Hudson. 2006. Using kinetic typography to convey emotion in text-based interpersonal communication. In *Proc. DIS '06*, DIS '06, pages 41–49, New York, NY, USA. ACM.

G. Leshed and J. Kaye. 2006. Understanding how bloggers feel: recognizing affect in blog posts. In *CHI'06 EA*.

H. Liu, T. Selker, and H. Lieberman. 2003. Visualizing the affective structure of a text document. In *CHI'03 EA*.

T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119.

P. Nakov, A. Ritter, S. Rosenthal, F. Sebastiani, and V. Stoyanov. 2016. Semeval-2016 task 4: Sentiment analysis in twitter. In *Proceedings of the 10th international workshop on semantic evaluation (SemEval'16)*.

R. Plutchik. 1980. *Emotion: A psychoevolutionary synthesis*. Harpercollins College Division.

S. Poria, A. Gelbukh, E. Cambria, A. Hussain, and G.-B. Huang. 2014. Emosenticspace: A novel framework for affective common-sense reasoning. *Knowledge-Based Systems*, 69:108–123.

J. A. Sánchez, I. Kirschning, J. C. Palacio, and Y. Ostróvskaya. 2005. Towards mood-oriented interfaces for synchronous interaction. In *Proc. CLIHC '05*, CLIHC '05, New York, NY, USA. ACM.

J. A. Sánchez, N. P. Hernández, J. C. Penagos, and Y. Ostróvskaya. 2006. Conveying mood and emotion in instant messaging by using a two-dimensional model for affective states. In *Proceedings of VII Brazilian Symposium on Human Factors in Computing Systems*, IHC '06, pages 66–72, New York, NY, USA, November. ACM.

D. Tang, F. Wei, N. Yang, M. Zhou, T. Liu, and B. Qin. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL (1)*, pages 1555–1565.

D. Tsetserukou, A. Neviarouskaya, H. Prendinger, N. Kawakami, M. Ishizuka, and S. Tachi. 2009. ifeel_im! emotion enhancing garment for communication in affect sensitive instant messenger. In *Human Interface and the Management of Information. Designing Information Environments*. Springer.

G. Wilson, D. Dobrev, and S. A. Brewster. 2016. Hot under the collar: Mapping thermal feedback to dimensional models of emotion. In *Proc. CHI'16*, CHI'16, pages 4838–4849, New York, NY, USA. ACM.

C.-H. Wu, Z.-J. Chuang, and Y.-C. Lin. 2006. Emotion recognition from text using semantic labels and separable mixture models. *ACM transactions on Asian language information processing (TALIP)*, 5(2).

R. Xu, T. Chen, Y. Xia, Q. Lu, B. Liu, and X. Wang. 2015. Word embedding composition for data imbalances in sentiment and emotion classification. *Cognitive Computation*, 7(2):226–240.

Y.-H. Yang and J.-Y. Liu. 2013. Quantitative study of music listening behavior in a social and affective context. *IEEE Transactions on Multimedia*, 15(6).

Z. Yeo. 2008. Emotional instant messaging with kim. In *CHI'08 EA*, CHI EA '08, April.