Help Me Learn! : Disparities in the distribution of Languages in the Livemocha Community

Jayant Venkatanathan
Departamento de Matemática e Engenharias
Universidade da Madeira
Funchal, Portugal
vjayant@m-iti.org

ABSTRACT
We study data collected from Livemocha - an online language learning social network - to identify speakers of those languages who might not sufficiently benefit from the community in their efforts to learn a language. We observe that speakers of language *A* that are learning language *B* are more likely to assist speakers of language *B* learning *A*. Based on this observation, we identify those pairs of languages *A* and *B* for which the likelihood for a speaker of *A* who is learning *B* of receiving sufficient assistance in her learning is low.

Author Keywords
Language learning, Contribution in Online social Networks

INTRODUCTION
A recent trend in the realm of online language learning portals has been the mushrooming of language learning social networking websites, where members share their expertise in the languages they know with the community, and can use the resources of the community to learn new languages. These sites are varied in the features that they offer. Some serve as a repository for lessons for various languages that members can contribute[1][2], some provide a means for teachers to register themselves so as to enable interested learners to contact them[3], while some serve as a platform for people learning various languages to interact and assist each other[4][5][6]. Livemocha[6] is one such online language learning social networking site which has gained a large number of users. Livemocha contains lessons for various language pairs. These lessons can contain exercises that include writing or reading out sentences in the language that the user is learning. Users who speak a certain language can check the exercise submissions of those trying to learn that language and provide feedback. Livemocha also provides a chat feature to allow instant messaging between users who are online and a message feature which allows users to send messages to each other.

Given that the primary motivation for the vast majority of users in an OSN like Livemocha is to learn or improve their skills in a language and not merely to contribute to the community, it would be reasonable to expect that a user *A* is more likely to provide feedback and help another user *B* if *B* speaks a language that *A* is trying to learn, so that *A* would benefit when *B* reciprocates the help. In such an environment, speakers of languages that not many members want to learn can be at a disadvantage. We investigate the presence of such language pairs where the number of users that learn the first and speak the second language is heavily disproportionate to the number of users who speak the second and learn the first.

RELATED WORK
The contribution of members in online social networks has been studied before in the context of various social networks such as facebook[7], wikipedia[8] and open source development teams[10]. [9] identifies characteristics of those posts in online forums that are most likely to receive feedback. A key distinction between our work from these is that (1) we attempt to study the beneficiaries of the contribution of the community and identify those users that are less likely to benefit from the contributions of the community, and (2) this being due their characteristics that they cannot be expected to change during the course of their learning, namely their languages of expertise.

METHOD
Each profile on Livemocha has a unique number corresponding to it which can be found on the url to the profile. These numbers varied from 1 (the first member) to a little over 4 million - the latest members - at the time of the collection of data. We used a pseudo-random number generator to generate about 26,000 random numbers between 1 and 4-million and crawled the profiles corresponding to these numbers. This was done to ensure that both old and new members of the site were fairly represented in the set profiles collected. These profile pages were then scraped to extract the languages that each user speaks and is learning. The website requires that each user specify the level of expertise with which she speaks (and correspondingly the level at which she is learning) a given language, and these levels fall into 5 classes : Beginner, Intermediate, Advanced, Fluent and Native. Since we were interested only in languages that a user could help another with learning, we considered that a user spoke a language only if she had selected Advanced, Fluent or Native as her level of expertise in the language. This data was then statistically analyzed.

RESULTS
The number of speakers and the number of learners for the top few languages with the highest numbers is shown tables 1 and 2.

Language	Speakers
Portuguese	10,675
English	8158
Spanish	6836
Mandarin Chinese	3731
Arabic	2564
Russian	1250
French	996
Turkish	824
Italian	676
Hindi	576
Korean	525
German	463
Farsi	364

Table 1. Number of Speakers

Portuguese topped the list for the number of speakers in our sample with 10,675 members, while there were 18,804 learners of English, making it the language that had the highest number of members learning. We calculated the ratio between the number of members learning each language and the number of members that spoke a language. Tables 3 and 4 summarizes this for a few Languages that have hight ratios.

LANGUAGE	LEARNERS
English	18,804
French	4920
Spanish	4710
German	2796
Italian	2421
Japanese	1894
Mandarin Chinese	949
Portuguese	883
Russian	773
Korean	567
Arabic	564
Hindi	293
Dutch	212

Table 2. Number of Learners.

Language	Speakers	Learners	S/L Ratio
Burmese	38	1	38.0
Indonesian	221	8	27.63
Portuguese	10675	883	12.09
Vietnamese	250	31	8.07
Urdu	208	27	7.70
Tagalog	166	24	6.92
Romanian	248	47	5.28
Filipino	126	13	9.69
Tamil	179	20	8.95
Arabic	2564	564	4.55

Table 3. Languages with high Speakers/Learners Ratios

Finally, for every pair of languages A and B, we calculated the ratio between the number of speakers of A learning B and the number of speakers of B learning A. We found ratios ranging from a very high (102.3 for Portuguese-Japanese) to pairs that were close to being balanced (≈ 1). Table 5 summarizes this for a few high ratio pairs and Figure 1 is a visual representation of this in the form of a network of languages

Language	Leaners	Speakers	L/S Ratio
Icelandic	68	8	8.5
Japanese	1894	145	13.06
German	2796	463	6.04
French	4920	996	4.94
Norwegian	64	15	4.27
Italian	2421	676	3.58
Swedish	118	41	2.88
English	18804	8158	2.30
Dutch	212	93	2.28
Finnish	85	34	2.5

Table 4. Languages with high Learners/Speakers Ratios

Figure 1. Visual Representation of the dataset as a Network

Language A	Language B	Speaks A learns B	Speaks B learns A	Ratio
Portuguese	Japanese	307	3	102.3
Turkish	Italian	44	1	44.0
Portuguese	Italian	723	21	34.43
Tamil	Hindi	31	1	31.0
Urdu	Arabic	26	1	26.0
Portuguese	English	5772	265	21.78
Spanish	French	791	157	5.038
Turkish	English	459	48	9.56
English	German	945	102	9.265
Thai	German	9	1	9.0
Portuguese	Greek	27	0	–
Arabic	English	1647	210	7.84
Mandarin Chinese	Japanese	444	12	37.0
Mandarin Chinese	Korean	199	36	5.53
Russian	German	99	20	4.95

Table 5..Ratios between pairs of Languages

Discussion

Tables 3 and 4 indicate that certain language speakers are inherently at a disadvantage (Table 3) , while speakers of some languages could expect an "excess" of help from the community(Table 4). For example, for every learner of Portuguese, there are roughly 12 speakers of Portuguese, and similarly this ratio for Burmese is 38. This implies that the potential ability of such members to contribute to the community is limited due to the fact that their skills are in lesser demand. On the other hand, for every Japanese speaker there are about 13 learners of Japanese, which implies that the contribution of a Japanese speaker is far more valuable and crucial to the community. Conversely, this also implies that a Japanese speaker is more likely to receive assistance from the community for a language that she is trying to learn.

Table 5 reveals that certain language pairs in our sample are highly imbalanced. For example, for every 102 Portuguese speaking members learning Japanese, there is only one Japanese speaker learning Portuguese, and for about every nine English speaking members learning German, there is one German speaker learning English. Consequently, many Portuguese speakers learning Japanese and English speakers learning German are unlikely to receive sufficient feedback on their lessons, while Japanese speakers learning Portuguese and German speakers learning English are likely to be flooded with feedback on their exercise submissions.

CONCLUSION

We have shown that the distribution of the number of members across various languages in Livemocha is highly unbalanced, and this can limit the extent of benefit that members that speak certain languages can receive from the community. More precisely, members who speak and learn certain language pairs are highly likely to be at a disadvantage. Moreover, the language that a person speaks is practically an immutable characteristic of that person during the course of his learning a language, and hence a member cannot change this. Such members can hence benefit from explicit measures by the community to address this issue. We identify such language pairs and the extent to which these pairs are unbalanced.

A key drawback of the study is that we have not been able to formally verify using ground truth data that these members indeed receive a lesser than average amount of assistance from the community. This would be a natural next step in our future work in this study.

ACKNOWLEDGMENTS

The author would like to thank Professor Vassilis Kostakos and all the participants of the Social Web (2009) course at the University of Madeira for the discussions and feedback during the course.

REFERENCES

1. Open Language Archives Community : http://www.language-archives.org
2. Unilang : http://www.unilang.org
3. italki : http://www.italki.com
4. Busuu : http://www.busuu.com
5. My Language Exchange : http://www.mylanguageexchange.com
6. Livemocha :http://www.livemocha.com
7. Burke M., Marlow C., Lento T.(2009). Feed Me: Motivating Newcomer Contribution in Social Network Sites. *ACM CHI 2009*
8. Cosley, D., Frankowski, D., Terveen, L., & Riedl, J. (2007). Suggestbot: Using intelligent task routing to help people find work in wikipedia. In Proceedings of the 12th acm international conference on intelligent user interfaces. New York: ACM Press
9. Arguello,J., Butler, B. S., Joyce, L., Kraut, R., Ling, K. S., RosŽ, C. P., et al. (2006). Talk to me: Foundations for successful individual-group interactions in online communities. In CHI 2006: Proceedings of the ACM

conference on human-factors in computing systems (pp. 959 - 968). New York: ACM Press.

10. Roberts, J., Hann, I.-H., & Slaughter, S. (2006). Understanding the motivations, participation and performance of open source software developers: A longitudinal study of the apache projects. Management Science, 52(7), 984 - 999.